

UNIVERSIDADE FEDERAL DO PARANÁ

ANDRÉ LUIZ EMIDIO DE ABREU

BOOTSTRAP E MODELOS DE SUPPORT VECTOR MACHINE – SVM

CURITIBA

2016

ANDRÉ LUIZ EMIDIO DE ABREU

BOOTSTRAP E MODELOS DE SUPPORT VECTOR MACHINE – SVM

Tese apresentada como requisito parcial a obtenção do grau de Doutor em Métodos Numéricos pelo Programa de Pós-Graduação em Métodos Numéricos em Engenharia, dos Setores de Tecnologia e de Ciências Exatas, na Área de Concentração em Programação Matemática e na Linha de Pesquisa em Métodos Estatísticos Aplicados à Engenharia.

Orientador: Prof. Dr. Anselmo Chaves Neto

CURITIBA

2016



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
Setor CIÊNCIAS EXATAS
Programa de Pós Graduação em MÉTODOS NUMÉRICOS EM
ENGENHARIA
Código CAPES: 40001016030P0

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em MÉTODOS NUMÉRICOS EM ENGENHARIA da Universidade Federal do Paraná foram convocados para realizar a arguição da Tese de Doutorado de **ANDRE LUIZ EMIDIO DE ABREU**, intitulada: "**BOOTSTRAP E MODELOS DE SUPPORT VECTOR MACHINE - SVM**", após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua aprovação.

Curitiba, 28 de Março de 2016.

Prof ANSELMO CHAVES NETO (UFPR)
(Presidente da Banca Examinadora)

Prof JAIR MENDES MARQUES (UFPR)

Prof LEONARDO EMMENDORFER (FURG)

Prof LILIANA MADALENA GRAMANI (UFPR)

Prof SILVANA HEIDEMANN ROCHA (UTFPR)

Prof VOLMIR EUGÊNIO WILHELM (UFPR)

RESUMO

A tese apresenta duas metodologias novas para a fase de ajuste de modelos de Máquinas de Vetores Suporte (*Support Vector Machines* – SVM) aplicados tanto em problemas de classificação de padrões, quanto em problemas de regressão (*Support Vector Regression* – SVR), e uma metodologia de obtenção de margens alternativa as já existentes para modelos SVR. Para as metodologias propostas, são apresentadas aplicações na medicina tais como: na classificação de grau da doença Ceratocone que atinge a córnea, na previsão do grau de astigmatismo com base nos exames clínicos. Apresenta, também, aplicação na climatologia, na previsão da temperatura e aplicação na parte hídrica quanto à previsão da evaporação de água em reservatórios. O primeiro método se baseia na criação de uma constante reguladora para as variáveis independentes dos modelos, levando em consideração o valor da correlação das mesmas com a variável dependente. Já o segundo método, utiliza a técnica computacionalmente intensiva de reamostragens *Bootstrap* para a fase de ajuste dos modelos SVM, liberando observações originais para a fase de teste. O *Bootstrap* é usado, ainda, na criação de novas margens flexíveis para os modelos SVR, com base no intervalo de confiança *bootstrap-t*, gerando uma forma alternativa de se obter as margens flexíveis. As soluções das modelagens foram obtidas pelo método *Least Squares Support Vector Machine* – LS-SVM, devido a sua menor complexidade matemática e computacional, em relação à solução dos modelos SVM. Em todas as aplicações, as metodologias desenvolvidas neste trabalho obtiveram resultados para a raiz do erro médio quadrático (*RMSE*) menores que para os modelos SVM convencionais, evidenciando assim a eficiência dos métodos desenvolvidos.

Palavras-Chave: *Bootstrap*, *Support Vector Machine*, Classificação, Regressão.

ABSTRACT

The thesis presents two new methodologies for adjust phase for models based on Support Vector Machines (Support Vector Machines - SVM) applied both in pattern classification problems, as in regression problems (Support Vector Regression - SVR), and a methodology of obtaining alternative margins existing for SVR models. The paper presents applications in medicine: classification of Keratoconus disease degree that affects the cornea, in predicting the degree of astigmatism based on clinical examinations; weather: on the forecast of temperature and finally in with forecast of evaporation of water in reservoirs. The first method is based on the creation of a regulatory constant for the independent variables of the model, taking into account the value of their correlation with the dependent variable. The second method uses the Bootstrap resampling for the training phase of the SVM models, releasing the original samples for testing phase. The third is to use the Bootstrap method to create the new flexible margins for SVR models, based on the bootstrap-t confidence interval, generating an alternative way to obtain the flexible margins. The solutions of the moldings were obtained by the least squares method Support Vector Machine - LS-SVM due to its smaller mathematical and computational complexity compared to the solution of SVM models. In all applications, the methodologies developed in this thesis obtained results for mean square error (RMSE) lower than for conventional SVM models, thus demonstrating the effectiveness of the methods developed.

Key-Words: Bootstrap, Support Vector Machine, Classification, Regression.

LISTA DE FIGURAS

FIGURA 1 – PREVISÕES DO NÍVEL DE EVAPORAÇÃO DE UM RESERVATÓRIO DE ÁGUA	15
FIGURA 2 – REALIZAÇÕES DE UM PROCESSO ESTOCÁSTICO	17
FIGURA 3 – ILUSTRAÇÃO DE CONJUNTOS LINEARMENTE SEPARÁVEIS.....	22
FIGURA 4 – POSSIBILIDADES DE HIPERPLANOS SEPARADORES.....	24
FIGURA 5 – EXEMPLOS DE MARGENS DE SEPARAÇÃO	25
FIGURA 6 – SEPARAÇÃO LINEAR INEFICIENTE PARA DOIS CONJUNTOS DISTINTOS	27
FIGURA 7 – HIPERPLANO ÓTIMO PARA A APLICAÇÃO DO SVR.....	32
FIGURA 8 – FUNÇÃO PERDA ε – <i>INSENSITIVE</i>	34
FIGURA 9 – EXEMPLO DE ERROS MÍNIMOS EM RELAÇÃO À FUNÇÃO DE REGRESSÃO.....	37
FIGURA 10 – MAPEAMENTO REALIZADO PELA FUNÇÃO <i>KERNEL</i>	40
FIGURA 11 – ALGORITMO DA DISTRIBUIÇÃO <i>BOOTSTRAP</i> DA ESTATÍSTICA $T_n(\underline{x}, F)$	46
FIGURA 12 – EXEMPLO DE APLICAÇÃO DAS MARGENS <i>BOOTSTRAP</i> PARA O MÉTODO SVR	58
FIGURA 13 – ALGORITMO DO TREINAMENTO <i>BOOTSTRAP</i> PARA O MÉTODO SVM.....	61
FIGURA 14 – ALGORITMO DO TREINAMENTO <i>BOOTSTRAP</i> PARA O MÉTODO SVR	63
FIGURA 15 – ALGORITMO DAS MARGENS FLEXÍVEIS <i>BOOTSTRAP</i> PARA O MODELO SVR.....	65
FIGURA 16 – INTERVALO DE VARIAÇÃO PARA AS PREVISÕES FEITAS PELO SVR <i>BOOTSTRAP</i>	66
FIGURA 17 – TREINAMENTO COM A INTRODUÇÃO DO MULTIPLICADOR DE CORRELAÇÃO	68
FIGURA 18 – OLHO AFETADO PELA CERATOCONE (a) E OLHO NORMAL (b) ..	70
FIGURA 19 – DIFERENÇA ENTRE OLHO NORMAL E OLHO COM ASTIGMATISMO	74

FIGURA 20 – COMPARAÇÃO GRÁFICA ENTRE AS PREVISÕES E OS VALORES REAIS.....	89
FIGURA 21 – MARGENS FLEXÍVEIS <i>BOOTSTRAP</i> PARA O MODELO LS-SVR...	89
FIGURA 22 – PREVISÕES PARA OS SEIS PACIENTES.....	92
FIGURA 23 – PREVISÕES E VALORES REAIS PARA OS SEIS PACIENTES.....	92
FIGURA 24 – MARGENS FLEXÍVEIS <i>BOOTSTRAP</i> PARA AS PREVISÕES.....	93
FIGURA 25 – REGIÃO DA ESTAÇÃO DE COLETA DOS DADOS	94
FIGURA 26 – PREVISÃO E MARGENS <i>BOOTSTRAP</i> DA TEMPERATURA	97
FIGURA 27 – COMPARAÇÃO ENTRE SVR E SVR <i>BOOTSTRAP</i>	98
FIGURA 28 – PREVISÕES DA COTAÇÃO DO DOLAR.....	100
FIGURA 29 – MARGENS FLEXÍVEIS PARA COTAÇÃO DO DOLAR.....	102

LISTA DE TABELAS

TABELA 1 – FICHA DOS EXAMES (CERATOCONES).....	71
TABELA 2 – FICHA DOS EXAMES (ASTIGMATISMO)	73
TABELA 3 – CLASSIFICAÇÃO ENTRE GRUPOS (1 E 2 CONTRA 3 E 4)	80
TABELA 4 – CORRELAÇÕES ENTRE AS VARIÁVEIS (CERATOCONES).....	81
TABELA 5 – RESUMO DOS PARÂMETROS E ERROS (1 E 2 CONTRA 3 E 4).....	81
TABELA 6 – CORRELAÇÕES PARA OS DADOS DE GRAU 1 E 2	82
TABELA 7 – CLASSIFICAÇÕES PARA PACIENTES COM GRAU 1 OU 2.....	83
TABELA 8 – RESUMO DOS PARÂMETROS E ERROS (GRAU 1 CONTRA 2)	84
TABELA 9 – MÍNIMO E MÁXIMO DAS VARIÁVEIS INDEPENDENTES GRAU 3....	85
TABELA 10 – MÍNIMO E MÁXIMO DAS VARIÁVEIS INDEPENDENTES GRAU 4..	85
TABELA 11 – VALORES OBTIDOS DAS CLASSIFICAÇÕES (1 CONTRA 2)	86
TABELA 12 – CORRELAÇÕES ENTRE AS VARIÁVEIS (EVAPORAÇÃO)	88
TABELA 13 – MÉTODO LS-SVR <i>BOOTSTRAP</i> PARA MARGENS FLEXÍVEIS	88
TABELA 14 – RESUMO DOS PARÂMETROS E ERROS (EVAPORAÇÃO).....	90
TABELA 15 – CORRELAÇÕES ENTRE AS VARIÁVEIS (ASTIGMATISMO).....	91
TABELA 16 – COMPARAÇÃO MÉTODOS SVR E SVR CORRELACIONADO.....	91
TABELA 17 – MARGENS FLEXÍVEIS <i>BOOTSTRAP</i> PARA O SVR.....	91
TABELA 18 – RESUMO DOS PARÂMETROS E ERROS (ASTIGMATISMO)	93
TABELA 19 – CORRELAÇÕES ENTRE AS VARIÁVEIS (TEMPERATURA)	95
TABELA 20 – CARACTERÍSTICAS PARA O TREINAMENTO <i>BOOTSTRAP</i>	96
TABELA 21 – MARGENS FLEXÍVEIS <i>BOOTSTRAP</i> PARA O SVR.....	96
TABELA 22 – VALORES CALCULADOS PARA AS 5 PREVISÕES SEMANAIS	97
TABELA 23 – RESUMO DOS PARÂMETROS E ERROS (TEMPERATURA).....	99
TABELA 24 – CARACTERÍSTICAS DO TREINAMENTO <i>BOOTSTRAP</i> (DÓLAR) ..	99
TABELA 25 – PREVISÕES PARA A COTAÇÃO DO DÓLAR	100
TABELA 26 – MARGENS FLEXÍVEIS PARA A COTAÇÃO DO DÓLAR	101
TABELA 27 – RESUMO DOS PARÂMETROS E ERROS (DÓLAR)	102

SUMÁRIO

1 INTRODUÇÃO	10
1.1 OBJETIVO	12
1.2 JUSTIFICATIVA	12
1.3 ESTRUTURA DA TESE	13
2 FUNDAMENTAÇÃO TEÓRICA	15
2.1 SÉRIES TEMPORAIS	15
2.2 MODELOS PARA SÉRIES TEMPORAIS	16
2.3 MODELOS DE PREVISÃO DE SÉRIES TEMPORAIS	18
2.3.1 Metodologia Box & Jenkins	18
2.4 APRENDIZADO DE MÁQUINA	20
2.5 <i>SUPPORT VECTOR MACHINE</i> (SVM)	21
2.5.1 Modelagem primal com as margens rígidas	21
2.5.2 Modelagem primal com margens flexíveis	26
2.5.3 Modelo Dual	28
2.6 <i>SUPPORT VECTOR REGRESSION</i> (SVR)	32
2.7 <i>LEAST SQUARES SUPPORT VECTOR MACHINES</i> (LS-SVM)	36
2.8 FUNÇÕES <i>KERNEL</i>	39
2.8.1 <i>Kernel</i> aplicado ao modelos LS-SVM	43
2.9 MÉTODO <i>BOOTSTRAP</i>	44
2.10 FUNÇÃO DISCRIMINANTE LINEAR DE FISHER	47
3 REVISÃO DE LITERATURA	49
3.1 BIOMETRIA	49
3.1.1 Classificação de faces e assinaturas	49
3.2 FINANÇAS	50
3.2.1 Instituições de crédito	50
3.2.2 Previsões de séries temporais financeiras	50
3.3 MEDICINA	51
3.4 INDÚSTRIA	51
3.4.1 Previsão de demanda de energia elétrica	51
3.4.2 Quantificar os adulterantes no leite em pó	51
3.5 PREVISÃO DE EVAPORAÇÃO DE ÁGUA EM RESERVATÓRIOS	52
4 MATERIAL E MÉTODO	53

4.1 CONFIGURAÇÃO DOS CONJUNTOS DE DADOS	53
4.1.1 Determinação dos parâmetros C e σ	54
4.2 MEDIDAS DE AVALIAÇÃO	55
4.2.1 Seleção de modelos	55
4.2.2 Avaliação das previsões	56
4.3 ERRO PADRÃO <i>BOOTSTRAP</i>	57
4.4 INTERVALO DE CONFIANÇA <i>BOOTSTRAP</i> – t	57
4.5 MARGENS <i>BOOTSTRAP</i>	58
4.6 METODOLOGIAS PROPOSTAS	59
4.6.1 Treinamento <i>Bootstrap</i> para os modelos SVM	59
4.6.2 Treinamento <i>Bootstrap</i> para os modelos SVR	62
4.6.3 Margens flexíveis <i>Bootstrap</i> para modelos SVR	64
4.6.4 SVM e SVR Correlacionado	66
4.7 DADOS DAS APLICAÇÕES REAIS	68
4.7.1 Amostras <i>Bootstrap</i>	68
4.7.2 Dados dos exames da Ceratocone	69
4.7.3 Dados do reservatório de Manasgaon	72
4.7.4 Dados do exame do grau de astigmatismo	72
4.7.5 Dados meteorológicos	74
4.7.6 Cotação do Dolar	76
5 RESULTADOS	78
5.1 CERATOCONE	78
5.1.1 Separação entre os graus (1, 2) e (3, 4) pelo método SVM Correlacionado	79
5.1.2 Separação entre graus 1 e 2 pelo método SVM Correlacionado	82
5.1.3 Separação entre graus 3 e 4 pelo método SVM <i>Bootstrap</i>	84
5.2 RESERVATÓRIO DE MANASGAON	87
5.2.1 Previsão de evaporação de água pelo método SVR Correlacionado	87
5.3 ASTIGMATISMO	90
5.3.1 Previsão do grau de astigmatismo pelo método SVR <i>Bootstrap</i>	90
5.4 PREVISÃO DA TEMPERATURA	94
5.5 COTAÇÃO DO DOLAR	99
6 CONCLUSÃO	103
REFERÊNCIAS	106

ANEXO 1 – DADOS COLETADOS PELOS EXAMES PARA A DOENÇA	
CERATOCONE (FELICIANO DE DEUS, 2015).....	111
ANEXO 2 – DADOS DO RESERVATÓRIO DE AGUA MANASGAON	
(DESWAL, PAL, 2008)	113
ANEXO 3 – VALORES DOS EXAMES DE ASTIGMATISMO	
(MAZUZE E MENDES, 2014).....	114
ANEXO 4 – VALORES CLIMÁTICOS PARA A PREVISÃO DA TEMPERATURA	
(REDE, 2016)	115

1 INTRODUÇÃO

A possibilidade de efetuar previsões vem se tornando a cada dia importante para a sociedade. Sejam elas meteorológicas, econômicas, sociais, entre outras. Os benefícios das previsões são inumeráveis: é capaz de ajudar um produtor rural a melhorar ou aumentar a sua produção agrícola; ajudar a evitar tragédias possibilitando salvar vidas quando seguidas com seriedade.

Pode-se citar como exemplo, as inúmeras tragédias causadas devido às chuvas torrenciais na região serrana do estado do Rio de Janeiro, onde se sabe historicamente que a cidade é atingida sazonalmente nos meses de verão, por fortes chuvas. A análise das séries temporais dos níveis de volume de chuva pode ter os resultados utilizados para justificar obras ambientais que a fim de evitar maiores problemas e diminuir ou evitar perdas de vidas.

A ideia de previsão trata desde previsões diárias do clima, nível de chuvas, chegadas de voos ou de ônibus em certo aeroporto ou terminal rodoviário, crescimento do PIB (Produto Interno Bruto) ou da taxa de inflação mensal de um país. Assim, este tema é útil e utilizado no dia a dia. Outro exemplo é a previsão de vazão de rios que deságuam em um lago represado de uma usina hidrelétrica. O conhecimento do volume de água é importante no dimensionamento da usina.

Além da previsão, é possível citar problemas de classificações ou identificações de padrões, utilizados, por exemplo, na medicina, para classificar se certo paciente possui um tumor maligno ou benigno, informando qual a urgência ou tipo de tratamento a que o paciente deve se submeter para obter o melhor resultado.

Desenvolver técnicas que possam elaborar previsões ou classificar padrões é de utilidade pública, sendo sempre necessário o advento de técnicas robustas e que sejam generalizáveis do ponto de vista técnico. Além da robustez e poder de generalização, é desejável que o tempo despendido para a obtenção das previsões e classificações seja razoável, uma vez que para uma pessoa com certo tipo de doença o tempo de diagnóstico pode ser crucial para o início do tratamento.

Assim, diversas técnicas foram desenvolvidas e produziram modelos de previsões e classificações para diversos tipos de fenômenos. Uma importante área da Estatística que estuda a construção de tais modelos de previsão é a da Análise de Séries Temporais, que a partir de séries temporais conhecidas ou sequências de valores que variam no tempo, se estudam e criam modelos estocásticos para prever

valores ou uma sequência de valores, a partir dos dados disponíveis, à frente no tempo.

É possível citar algumas técnicas desenvolvidas para o estudo de previsões de séries temporais, tais como: os métodos automáticos ou caixa preta (década de 1960), modelos Box & Jenkins (1970) e para a era pós Box & Jenkins, os modelos do filtro adaptativo, método Forsys, método ARARMA, combinações de previsões, métodos ARCH e GARCH, técnicas de redes neurais e algoritmos genéticos (CHAVES NETO, 2009).

Além das áreas citadas, outra área que vem recebendo crescente atenção é a do estudo de técnicas de reconhecimento de padrões. Nessa linha, a técnica de aprendizado de máquinas conhecida por SVM (*Support Vector Machine* – Máquina de Vetor de Suporte) vem ganhando grande destaque, sendo utilizada para reconhecimento de padrões, e os resultados obtidos são superiores aos alcançados por outras técnicas de aprendizado em várias aplicações (LORENA; CARVALHO, 2003).

A ideia inicial de criar algoritmos para reconhecer padrões surgiu através de Fisher em 1936. Mais tarde, em 1962, Rosenblatt (JONHSON, 1988) sugeriu o algoritmo *Perceptron* de aprendizagem, que é à base das Redes Neurais.

Em 1965, Vladimir Vapnik propôs um algoritmo para encontrar um hiperplano separador em uma classificação linear (VAPNIK; CHERVONENKIS, 1974). O objetivo inicial era encontrar um hiperplano separador que tivesse a máxima margem, ou seja, a maior distância entre os vetores suportes, usando a metodologia de aprendizagem. Entretanto, em 1992, Bernhard Boser, Isabelle Guyon e Vapnik criaram uma classificação não linear, aplicando as Funções *Kernel* (BOSER, GUYON E VAPNIK, 1992).

Em 1995, Corinna Cortes e Vapnik sugeriram uma modificação no algoritmo, introduzindo variáveis de folgas para obter uma margem suave no hiperplano separador, permitindo a classificação incorreta dos dados, chamado de modelo de margens flexíveis (CORTES; VAPNIK, 1995).

As buscas por métodos melhores não cessam. Foram criadas algumas variações para o SVM, sendo ideais em certas condições sobre o conjunto de dados, por exemplo, número elevado de pontos. Neste trabalho também será estudada a variante utilizada para problemas de regressão, chamado de *Support Vector Regression* - SVR (Vetor Suporte para Regressão).

1.1 OBJETIVO

Este estudo visou elaborar uma metodologia alternativa às metodologias já existentes e consolidadas, na literatura, para o ajuste do método *Support Vector Machine* (SVM), tanto para os problemas de classificação de padrões, quanto para problemas de regressão. Ainda, procurou melhorar os resultados para cada tipo de aplicação, ou seja, procurou diminuir o erro em cada classificação ou previsão. Além disso, buscou verificar a generalização para diversos tipos de problemas, tanto em previsões de séries temporais, quanto na classificação de padrões e reconhecimento de diferentes classes a serem rotuladas.

Por se tratar de uma metodologia computacional, uma característica desejável da SVM é um alto grau de robustez, ou seja, que pudesse ser aplicada a uma ampla gama de problemas, não se restringindo apenas a séries, ou problemas teóricos, e sim que fosse possível de ser generalizada para utilização em maior número de padrões de séries reais, além disso, havia a necessidade de que o tempo computacional seja razoavelmente baixo.

Ao final da aplicação das metodologias propostas, os resultados foram confrontados com os resultados obtidos pelas metodologias já existentes e consolidadas na literatura referenciada, com a finalidade de verificar, o seu desempenho e sua validade frente aos métodos já existentes.

Assim, o objetivo almejado ao fim deste estudo, foi poder contribuir com uma metodologia robusta, capaz de se adaptar ao contexto estudado, que pode ser utilizada, tanto na área acadêmica, como no cotidiano.

1.2 JUSTIFICATIVA

Métodos robustos de classificação de padrões e de previsão de séries possuem certa importância nas análises estatísticas. Quanto melhor a classificação ou previsão, maior será o benefício, diminuindo danos causados por catástrofes, informando possíveis épocas de seca, diminuição ou aumento de produtividade, diminuindo os erros de prognósticos médicos, informando o possível comportamento da economia, entre outras informações. Assim, a elaboração de uma técnica robusta se faz necessário, porém, tal técnica deve ter no mínimo um desempenho

semelhante às já existentes, a fim de gerar uma metodologia ao menos alternativa em equivalência às demais.

Segundo Lorena e Carvalho (2003), os resultados da aplicação da técnica *Support Vector Machine* são comparáveis aos obtidos por outros algoritmos de aprendizado, como as Redes Neurais Artificiais (RNA's) (HAYKIN, 1999). Em alguns casos, esses resultados têm se mostrado superiores, tal como na detecção de faces em imagens (HEARST *et al.*, 1998), na categorização de textos (HEARST *et al.*, 1998) e em aplicações em Bioinformática (ZIEN *et al.*, 2000).

Ainda, as técnicas de Aprendizado de Máquina (AP), onde o SVM se enquadra, são indicadas quando se tem um grande volume de dados, com alta complexidade ou alta dimensionalidade, e ainda, onde os métodos clássicos não se demonstram eficientes (GONÇALVES, 2008).

A técnica das Redes Neurais Artificiais e a da Programação Genética já são utilizadas rotineiramente em problemas de previsão (BUZZI, 2012; ALMEIDA, 2010; GEVERT, 2009), e tais aplicações foram consideradas bem sucedidas, pois, no mínimo, seus resultados se equiparam aos resultados das técnicas já consolidadas.

Além disso, devido às características vantajosas existentes, tais como: sua boa generalização, robustez em grandes dimensões e teoria bem definida, a metodologia do *Support Vector Machine* foi escolhida para este estudo, visando uma melhoria nas técnicas já existentes, e sua escolha foi feita com base no desempenho elevado apresentado na literatura referenciada.

O desenvolvimento da nova técnica, *Support Vector Machine Bootstrap*, se deu aos constantes problemas, vividos pelo autor, relacionados à falta de dados, a um número considerável, para as análises estatísticas, devido a problemas burocráticos, econômicos ou por necessitar de um intervalo de tempo elevado para a coleta dos dados.

1.3 ESTRUTURA DA TESE

Esta tese está organizada da seguinte forma. A primeira parte é composta pelos capítulos Introdução, Fundamentação Teórica e Revisão de Literatura. Já a segunda parte abrange os capítulos: Material e Métodos, Resultados e Discussão, Conclusão e as referências.

O capítulo 1, Introdução, contempla a apresentação do tema estudado, o objetivo e a justificativa. Já no capítulo 2, Fundamentação Teórica, é descrita a revisão bibliográfica, sobre os métodos utilizados no estudo, baseada nos livros clássicos do tema. São apresentadas as definições e formulações dos modelos de classificação (SVM) e de regressão (SVR) que foram aplicados aos dados de ajuste e teste. No capítulo 3, Revisão de Literatura, são abordados artigos referentes às técnicas mencionadas anteriormente, ou seja, na fundamentação teórica. O capítulo 4, aborda a descrição dos materiais e dos métodos utilizados no desenvolvimento do trabalho, bem como a formulação e sequência da implementação da metodologia. No capítulo 5, são apresentados os resultados encontrados e foi feita uma discussão sobre a metodologia desenvolvida, bem como as comparações entre os métodos abordados. Finalmente, o capítulo 6 apresenta a conclusão, as considerações finais e as referências utilizadas neste estudo.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 SÉRIES TEMPORAIS

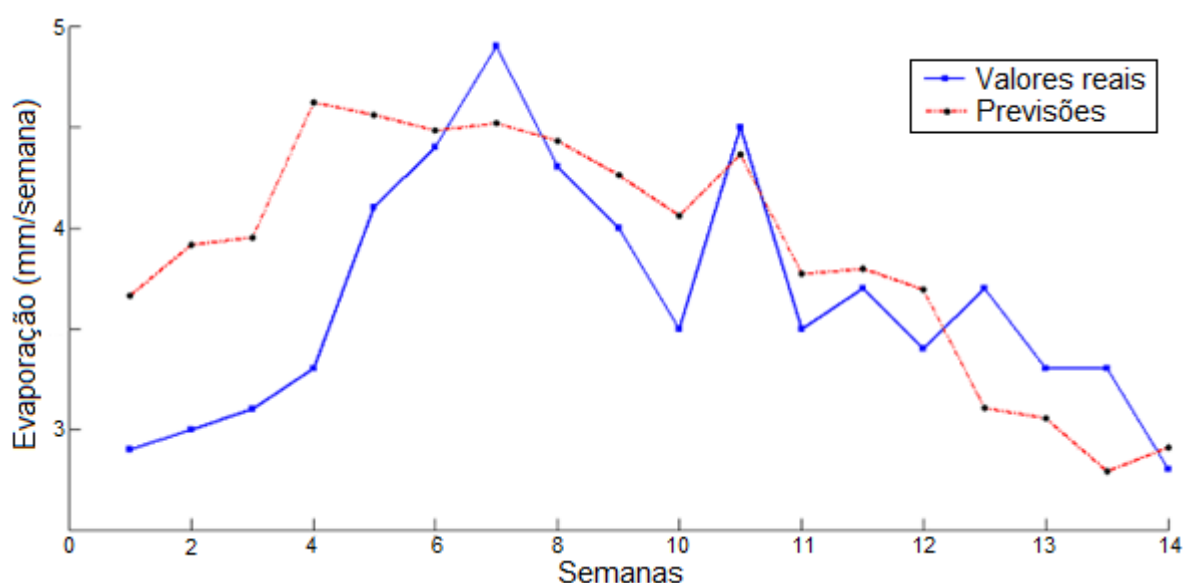
Uma série temporal pode ser caracterizada como um conjunto de observações ordenadas no tempo e podem ser classificadas como séries temporais discretas ou contínuas (MORETTIN; TOLOI, 1981).

A representação de uma série temporal pode ser feita da seguinte forma: por enumeração $\{Z_1, ..., Z_t, ..., Z_{n-1}, Z_n\}$ ou por compreensão $\{Z_t, t = 1, 2, ..., n\}$. Geralmente as observações da série são feitas no mesmo intervalo de tempo fixado, h .

Para este trabalho foi adotado o padrão apresentado a seguir para a notação de uma série temporal. De um modo geral, uma série temporal pode ser considerada como um vetor $\underline{Z}(t)$ de dimensão r , onde cada $Z(t_i)$ é uma observação e por sua vez, o parâmetro t é referido como o tempo. Mas uma série poderá ser função de algum outro parâmetro físico, como o espaço ou volume (MORETTIN; TOLOI, 2006).

A (FIGURA 1) apresenta o gráfico de duas séries temporais referentes às previsões da evaporação de água em um reservatório, durante 18 semanas seguidas (ABREU; CHAVES NETO, 2015).

FIGURA 1 – PREVISÕES DO NÍVEL DE EVAPORAÇÃO DE UM RESERVATÓRIO DE ÁGUA



FONTE: Adaptado de Abreu, 2015.

Segundo Morettin e Toloi (1981) ao se analisar uma série temporal surgem questões que são de certa relevância para o estudo, tais como:

- i. Investigar o mecanismo gerador da série (processo estocástico);
- ii. Fazer previsões de valores futuros da série, que podem ser de curto prazo, como séries de vendas, de produção ou de estoque, ou de longo prazo, como séries populacionais, de produtividade, etc.;
- iii. Apenas descrever o comportamento da série, buscando identificar padrões de comportamento tais como sazonalidade, tendências, entre outros.

É esperado que uma série deste tipo apresentasse a característica de estacionariedade, ou seja, se desenvolva no tempo aleatoriamente ao redor de um valor médio constante, refletindo alguma forma de equilíbrio estável e, também, tenha variância constante. Essas são condições necessárias para se estimar os parâmetros do modelo a serem ajustados às observações, quando se aplica a Metodologia Box & Jenkins. Porém, muitas vezes as séries temporais reais apresentam alguma forma de não estacionariedade (MORETTIN; TOLOI, 1981). Então, existem técnicas para tentar tornar estacionárias essas séries, tanto na média quanto na variância, aplicando diferenciação ou transformação de variável.

2.2 MODELOS PARA SÉRIES TEMPORAIS

Uma série temporal pode ser um processo estocástico especial, ou seja, um processo ergódico, onde se entende por processo ergódico processos onde todas as suas propriedades estatísticas podem ser determinadas a partir de uma única realização. Sendo assim, é possível estimar os parâmetros do processo gerador com uma única realização da série, supondo estacionariedade na média e na variância.

Os modelos utilizados para descrever esses processos são classificados e controlados por leis probabilísticas. Qualquer classificação feita para esses modelos, considera uma variedade muito grande de modelos diferentes para descrever o comportamento de uma série particular. A definição de um processo estocástico é dada por (MORETTIN; TOLOI, 1978):

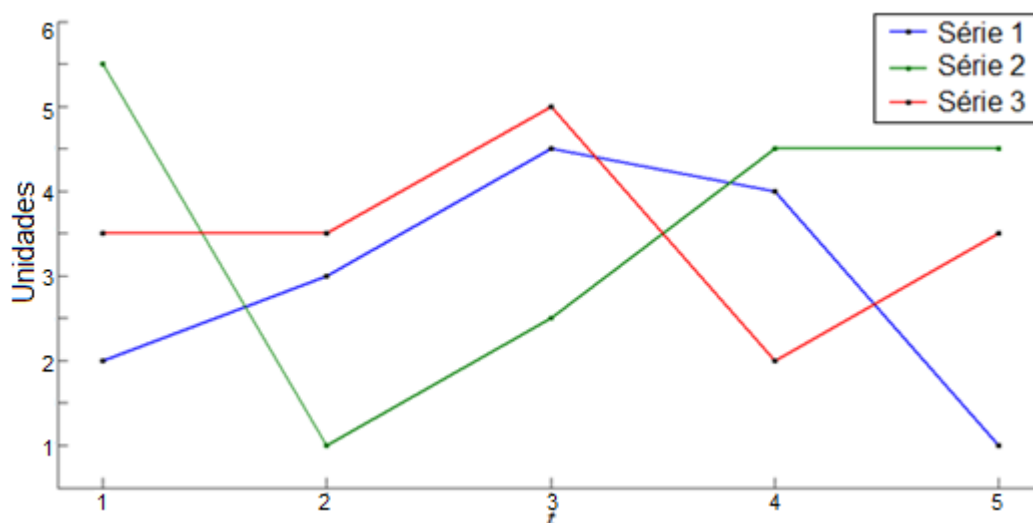
Definição: Seja T um conjunto arbitrário de índices. Um processo estocástico $Z(\omega, t)$ (função de duas variáveis ω e t) é uma família $Z = \{Z_t, t \in T\}$, tal que, para cada $t \in T$, Z_t é uma variável aleatória.

Logo, um processo estocástico é uma família de variáveis aleatórias (v. a.'s), que se supõem definidas no mesmo espaço de probabilidades (Ω, A, P) , em que Ω é o espaço amostral do experimento, A é a σ -Álgebra associada a Ω e P a medida de probabilidade. O conjunto T é normalmente tomado como o conjunto dos números inteiros Z . Desta forma processo estocástico é interpretado como uma família de trajetórias ou realizações para cada evento ω fixado, ou melhor, para cada $\omega \in \Omega$ fixado, obtém-se uma função de t , ou seja, uma trajetória ou realização do processo.

O conjunto $\{Z_t, t \in T\}$ é chamado Espaço dos Estados e os valores Z_t são chamados Estados. Dado o espaço de estados $\{Z_t, t \in T\}$ com T sendo um conjunto finito ou enumerável, o processo pode ser discreto ou contínuo, com base no conjunto T (CHAVES NETO, 2009).

A (FIGURA 2) apresenta um exemplo de três possíveis realizações de um processo estocástico. Observa-se que no período 1 têm-se cinco observações da variável aleatória Z_1 (Série 1), no período 2 tem-se as observações da v. a. Z_2 (Série 2) e no período 3 os registros de Z_3 (Série 3).

FIGURA 2 – REALIZAÇÕES DE UM PROCESSO ESTOCÁSTICO



FONTE: O autor (2016).

2.3 MODELOS DE PREVISÃO DE SÉRIES TEMPORAIS

Ao longo dos anos, diversos modelos de predição para séries temporais foram desenvolvidos, tais como os modelos de alisamento exponencial, auto regressão, filtragem adaptativa, os modelos Box & Jenkins, redes neurais, composição de previsões, modelos ARCH e GARCH, entre outros.

A grande maioria dos métodos de previsão encontrados na literatura se baseia na ideia de que as observações passadas contêm informações sobre o padrão de comportamento das séries temporais estudadas. Assim, o propósito dos métodos é distinguir o padrão encontrado de qualquer ruído que possa estar contido nas observações e então usar esse padrão para prever valores futuros da série.

Uma grande classe de métodos de previsão, que tenta tratar ambas as causas de flutuação em séries de tempo é a do alisamento exponencial. Técnicas específicas, desse tipo, assumem que os valores extremos da série representam a aleatoriedade e, assim, através do alisamento desses extremos, pode ser identificado o padrão básico. A grande popularidade atribuída aos métodos de alisamento é devida à simplicidade, à eficiência computacional e à sua razoável precisão (MORETTIN; TOLOI, 1981).

Segundo Morettin e Toloi, 1981, a análise de regressão foi por certo tempo um método aceitável para ajustar modelos autoregressivos, com o objetivo de calcular previsões. Porém, tal método não é útil quando se trabalha com um número pequeno de observações, pois a hipótese de independência dos resíduos é quase sempre violada, produzindo assim, estimadores inconsistentes.

2.3.1 Metodologia Box & Jenkins

A metodologia de Box e Jenkins foi um dos estudos responsável pelo grande desenvolvimento e a correspondente formalização da área de estudo de Séries Temporais. O trabalho dos pesquisadores Box e Jenkins foi baseado no importante resultado de Wold (1938) “qualquer série temporal pode ser representada por uma estrutura de médias móveis infinitas”, ou melhor, “qualquer processo estocástico estacionário Y_t pode ser representado como a soma de dois processos mutuamente

inter-relacionados, $Y_t = D_t + A_t$, onde D_t é linearmente determinístico (sistemático) e A_t é um processo Médias Móveis infinito ($MA(\infty)$)” (CHAVES NETO, 2009).

Segundo Morettin e Toloi, 1981, a metodologia de Box e Jenkins consiste em ajustar modelos autoregressivos integrados de médias móveis, da classe denominada $ARIMA(p, d, q)$ a um conjunto de dados já conhecidos.

A estratégia para a construção do modelo é baseada em um ciclo iterativo, onde a estrutura do modelo é baseada nos próprios dados. Os estágios do ciclo iterativo são, segundo Morettin e Toloi (2006), os seguintes:

- a) Uma classe geral de modelos é considerada para a análise (especificação);
- b) Há a identificação de um modelo, com base na análise de autocorrelação, autocorrelação parcial e outros critérios;
- c) A seguir, vem à fase de estimação, na qual os parâmetros do modelo identificado são estimados;
- d) Finalmente, há a verificação do modelo ajustado, através de uma análise de resíduos, para saber se ele é adequado.

Caso o modelo não seja adequado, inicia-se o ciclo novamente voltando à fase de identificação.

A forma geral de representar uma classe de modelos que podem ser ajustados é aquela composta por modelos lineares e é denominada por Modelos Auto-Regressivos Integrados Médias-Móveis de ordem (p, d, q) , ou seja, $ARIMA(p, d, q)$ onde p é o grau do polinômio $\phi(B)$, q o grau do polinômio $\theta(B)$ e d o grau de diferenciação ∇^d , ou seja, $\phi(B)\nabla^d Z_t = \theta(B)a_t$ onde $\nabla^d Z_t = \omega_t$ e desse modo tem-se $\phi(B)\omega_t = \theta(B)a_t$, que é um modelo ARMA (CHAVES NETO, 2009).

2.4 APRENDIZADO DE MÁQUINA

O Aprendizado de Máquina (AM) é um campo de pesquisa da Inteligência Computacional que estuda o desenvolvimento de métodos capazes de extrair conceitos (conhecimento) a partir de amostras de dados (MITCHELL, 1997).

Em geral, os diversos algoritmos de AM são utilizados de forma a gerar classificadores para um conjunto de exemplos. Neste contexto, classificação é conceituada como o processo de atribuir, a uma determinada informação, o rótulo da classe a qual ela pertence (RUSSEL; NORVIG, 1995). Portanto, as técnicas de AM são empregadas na indução (com base em um conjunto de treinamento) de um classificador, que deve ser capaz de prever, com o menor erro possível, a classe de instâncias quaisquer do domínio em que ele foi treinado. Para tanto, três métodos podem ser utilizados na geração de um preditor por meio de técnicas de Aprendizado de Máquinas: supervisionado, não supervisionado e por reforço (HAYKIN, 1999). A escolha de um método de aprendizado determina a maneira como o algoritmo de AM se relaciona com seu meio ambiente, ou seja, o modo como ocorrerá o seu aprendizado por meio de um conjunto de dados.

No método de aprendizado supervisionado tem-se a figura de um “professor externo”, o qual apresenta um conhecimento do ambiente representado por conjuntos de exemplos na forma entrada/saída. Nesse caso, o algoritmo de AM é treinado a partir de conjuntos de exemplos já rotulados com o objetivo de aprender uma função desejada (LORENA; CARVALHO, 2003).

No método de aprendizado não supervisionado não há a presença dos rótulos. O algoritmo de AM aprende a representar (ou agrupar) as entradas submetidas segundo uma medida de qualidade.

O método denominado aprendizado "por reforço", o aprendizado se dá por meio de recompensas ou não ao indutor, dependendo de seu desempenho em aproximar a função desejada.

O desenvolvimento das técnicas em AM abriu espaço para metodologias que fazem analogia aos sistemas biológicos, tais como as Redes Neurais Artificiais cujos conceitos são baseados nos mecanismos de aprendizado cerebrais e os Algoritmos Genéticos, que têm seus conceitos inspirados no processo de evolução natural e na genética. No campo de aprendizado simbólico têm-se as Árvores de

Decisão. Inspirado nos processos cognitivos houve o desenvolvimento dos Sistemas de Raciocínio Baseado em Casos (NETTO, 2005).

2.5 SUPPORT VECTOR MACHINE (SVM)

A ideia inicial de criar algoritmos para reconhecer padrões surgiu a partir de estudos realizados por Fisher em 1936, em 1962, Rosenblatt sugeriu o algoritmo *Perceptron* de aprendizagem. A fundamentação teórica do reconhecimento de padrões usando vetores suportes foi formalizada em 1992, após três décadas de pesquisa. O objetivo inicial era encontrar um hiperplano separador que tivesse a máxima margem, ou seja, a maior distância entre os vetores suportes, usando a metodologia de aprendizagem.

Em 1963, Vladimir Vapnik propôs o algoritmo do hiperplano de máxima margem, para uma classificação linear. As pesquisas continuaram com a intenção de encontrar uma separação não-linear. Em 1992, Bernhard Boser, Isabelle Guyon e Vladimir Vapnik sugeriam a separação não linear, utilizando as funções *Kernel*, que transformam o espaço de entrada em um espaço de dimensão maior, onde os dados são linearmente separáveis (BOSER *et al.*, 1992)

A forma original de modelar o problema do SVM é através do seu modelo primal, onde se pode determinar o hiperplano separador de padrões distintos. A seguir apresentam-se dois modelos de SVM, na forma dual. O primeiro enfatiza a máxima margem de classificação, caso linear e o segundo considera os conjuntos de treinamento não lineares.

2.5.1 Modelagem Primal com as Margens Rígidas

O SVM é uma metodologia que basicamente classifica os padrões em dois conjuntos, mas que pode ser ampliado para classificações em múltiplas classes. Devido à metodologia utilizar margens (vetores) de separação para os conjuntos, as quais são deslocamentos com distância igual a um, tem-se o conjunto de classe -1 e outro de classe +1.

Por se tratar de uma metodologia de aprendizado de máquinas, com aprendizagem supervisionada, é introduzida uma notação para referenciar o conjunto de treinamento, dados de entrada e saídas.

Seja o conjunto de treinamento:

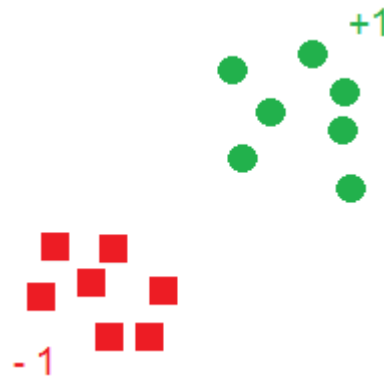
$$S = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \subseteq \{X \times Y\}^l, \quad (1)$$

com l o número de elementos no conjunto de treinamento e (x_i, y_i) são os pontos de treinamento, onde cada ponto é representado por (x_i, y_i) , para $i = 1, \dots, l$, onde $x_i \in \mathbb{R}^n$, $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$, denotando o valor de entrada pertencente ao espaço de entrada, que representa a quantificação das características, e $y_i \in \{-1, +1\}$ é a saída binária correspondente. Dado um ponto de entrada x , a saída do SVM é representada pelo sinal da função $f(x)$, ou seja, $\text{sig}(f(x))$.

O objetivo do modelo do SVM é encontrar um hiperplano ótimo, onde possa classificar corretamente todos os pontos de treinamento e, além disso, com a maior margem de separação possível para os vetores da classe -1 e da classe $+1$.

Considere inicialmente dois conjuntos linearmente separáveis tais como o ilustrado na (FIGURA 3).

FIGURA 3 – ILUSTRAÇÃO DE CONJUNTOS LINEARMENTE SEPARÁVEIS



FONTE: O autor (2016).

Para este exemplo, uma classificação linear pode ser obtida pela função real $f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ da seguinte forma: a entrada $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ é considerada de classe $+1$ se $f(x) \geq 0$ e, caso contrário, é considerada de classe -1 .

Assim, a função de decisão $f(x)$ é representada por:

$$f(x) = \underline{w}^t \underline{x} + b, \quad (2)$$

onde:

$\underline{w} \in \mathbb{R}^n$ é o vetor de pesos;

\underline{w}^t é o vetor transposto de \underline{w} ;

$b \in \mathbb{R}$ é chamado de *bias*.

Caso o espaço de características tenha dimensão dois, o hiperplano separador obtido é uma reta. Caso o espaço de características tenha dimensão n o hiperplano separador obtido é um hiperplano de dimensão $n-1$. O hiperplano separador divide o espaço de características em dois subespaços, sendo um denominado classe -1 e o outro classe $+1$.

A classificação de cada padrão x do conjunto de treinamento é dada conforme a proximidade em relação às margens do hiperplano separador, ou seja, será classificado como pertencente à classe -1 se estiver mais próximo da margem negativa, caso o valor da função de decisão seja negativa ($\underline{w}^t \underline{x} + b = -1$) e será pertencente a classe $+1$ se estiver mais próximo da margem positiva ($\underline{w}^t \underline{x} + b = +1$).

Um padrão x_i é considerado classificado corretamente se ele estiver fora da margem de separação de sua classe, ou seja, se $y_i = +1$ então deve satisfazer ($\underline{w}^t \underline{x} + b \geq +1$) e se $y_i = -1$ então deve satisfazer ($\underline{w}^t \underline{x} + b \leq -1$). Assim, é possível combinar as duas restrições em uma única só (BURGES, 1998):

$$y_i[\underline{w}^t \underline{x}_i + b] \geq +1, \quad i = 1, 2, \dots, l. \quad (3)$$

Denomina-se margem funcional do ponto (x_i, y_i) com relação ao hiperplano $f(x)$ a quantidade dada por:

$$\gamma_i = y_i \left(\left\langle \underline{w}^t \underline{x}_i \right\rangle + b \right). \quad (4)$$

se $\gamma_i > 0$, então o ponto está classificado corretamente.

A menor margem funcional é denominada margem funcional do hiperplano definida por γ , e, se estiver normalizada, denomina-se margem geométrica. Ou seja,

se w é um vetor unitário, a margem geométrica é igual à margem funcional (CRISTIANINI; SHAW-ETAL, 2000).

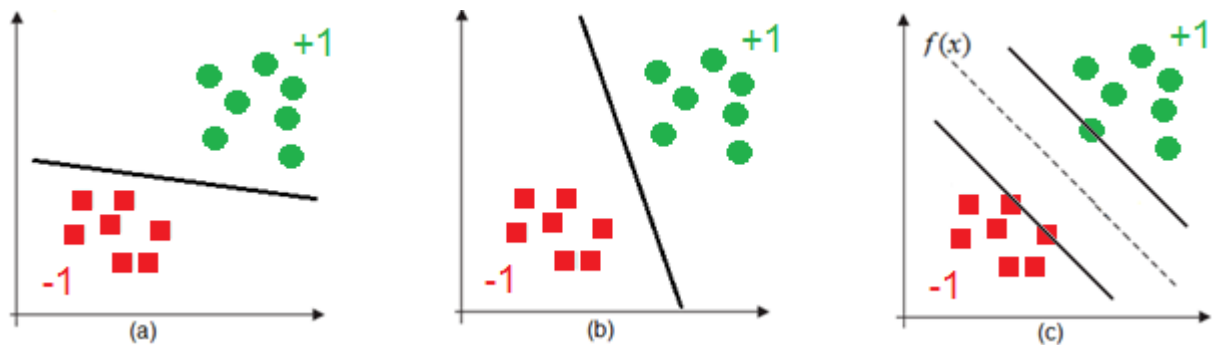
Fixado o valor $\gamma > 0$, pode-se definir a margem de folga ξ_i do ponto (x_i, y_i) em relação ao hiperplano separador como:

$$\xi_i = \max\left\{0, \gamma_i - y_i\left(\left\langle \underline{w}^t \underline{x}_i \right\rangle + b\right)\right\}. \quad (5)$$

A expressão (5) informa a quantidade em que o ponto falha para obter a margem funcional γ .

O objetivo é obter um hiperplano com a melhor generalização e robustez, dado que existem muitas possibilidades de separação dos conjuntos. A (FIGURA 4) apresenta três possibilidades de separação dos conjuntos de dados. Sendo que na (FIGURA 4) (a) e (b) são apresentadas separações quaisquer, sem a necessidade da máxima distância entre o hiperplano separador e os dados. Já para a (FIGURA 4c) está apresentada o gráfico do hiperplano separador ótimo.

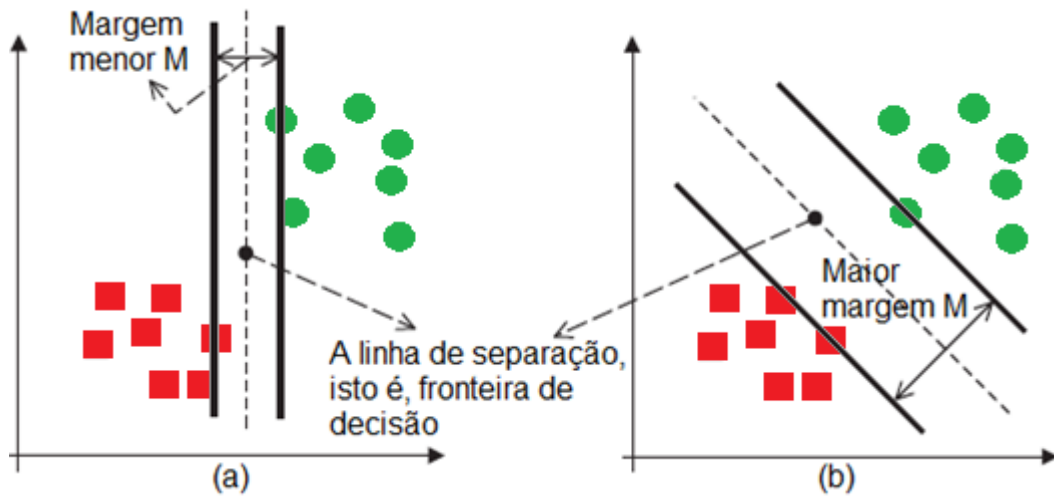
FIGURA 4 – POSSIBILIDADES DE HIPERPLANOS SEPARADORES



FONTE: O autor (2016).

O processo de treinamento do SVM consiste na obtenção de valores para os pesos w e do termo *bias* b de forma a maximizar a distância entre as margens (FIGURA 5b). Desta maneira, o SVM se torna robusto a pequenas variações no conjunto de treinamento, possibilitando uma melhor generalização.

FIGURA 5 – EXEMPLOS DE MARGENS DE SEPARAÇÃO



FONTE: Adaptado de Kecman (2005).

A (FIGURA 5) apresenta duas separações lineares perfeitas e também as margens. Para a (FIGURA 5a), a separação não possui ainda a margem máxima passível de ser calculada, ou seja, é considerada não-ótima, já a (FIGURA 5b) apresenta, para o mesmo conjunto de dados de treinamento, margem máxima possível para esses dados de treinamento, considerada uma separação ótima.

Considerando as margens do hiperplano de separação como $\underline{w}^t \underline{x} + b = +1$ e $\underline{w}^t \underline{x} + b = -1$, almeja-se obter a maior distância entre as margens para encontrar a melhor generalização e robustez do modelo do SVM. Essa afirmação é garantida pelo teorema a seguir:

Teorema: Caso os conjuntos sejam linearmente separáveis, têm-se que o hiperplano encontrado $\underline{w}^t \underline{x} + b = 0$ separa os dois conjuntos, cujas margens são definidas por:

$$\begin{aligned} \underline{w}^t \underline{x} + b &= +1, \\ \underline{w}^t \underline{x} + b &= -1. \end{aligned} \tag{6}$$

A distância entre as margens é dada por $d = \frac{2}{\|\underline{w}\|}$, logo, para obter o melhor hiperplano, ou seja, aquele hiperplano que está mais longe dos dois conjuntos, deve-se maximizar $d = \frac{2}{\|\underline{w}\|}$, ou de forma equivalente, minimizar $\frac{1}{2} \underline{w}^t \underline{w}$.

A maximização da margem de separação entre as classes, quando os conjuntos são linearmente separáveis, com o objetivo de obter um hiperplano ótimo, através do SVM, é resolvida através de um problema de otimização restrito. O problema primal do SVM é dado por:

$$\begin{aligned} \min \quad & \frac{1}{2} \underline{w}^t \underline{w} \\ \text{s.a} \quad & \underline{w}^t \underline{x} + b \leq -1, \quad x \in -1 \\ & \underline{w}^t \underline{x} + b \geq +1, \quad x \in +1, \end{aligned} \quad (7)$$

ou de forma reduzida:

$$\begin{aligned} \min \quad & \frac{1}{2} \underline{w}^t \underline{w} \\ \text{s.a} \quad & y_i (\underline{w}^t \underline{x} + b) \leq +1, \end{aligned} \quad (8)$$

onde:

\underline{w} e b são variáveis com $\underline{w} \in \mathbb{R}^n$ e $b \in \mathbb{R}$.

Tal classificador recebe o nome de *Perceptron*, para redes neurais, e Discriminante Linear, para estatística, onde o vetor \underline{w} é referenciado como vetor de pesos e b conhecido por limiar ou *bias*.

2.5.2 Modelagem Primal com Margens Flexíveis

Por serem raros os casos onde os dados podem ser facilmente separáveis linearmente desenvolveu-se a modelagem com margens flexíveis.

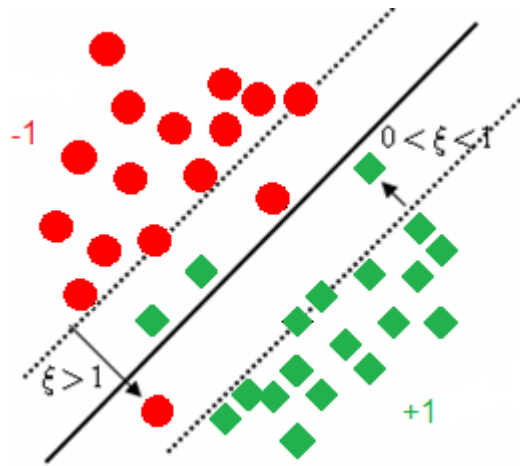
Assim, dado um conjunto de treinamento S não separável linearmente, o problema primal apresentado anteriormente pode separar padrões, porém, classificará alguns padrões incorretamente modificando seu valor na função de decisão.

Para impedir tal situação e, também, aumentando a capacidade de generalização do SVM acrescentam-se variáveis de folga, $\xi_i \geq 0$, associadas para cada vetor de treinamento x_i , sendo acrescida na restrição do SVM como:

$$y_i (\underline{w}^t \underline{x} + b) \geq +1 - \xi_i. \quad (9)$$

As variáveis de folga são obtidas durante o processo de treinamento do SVM. Quando um padrão x_i está separado corretamente tem-se $\xi_i = 0$. Caso um padrão x_i esteja entre a margem de separação de sua classe e o hiperplano separador, tem-se $0 < \xi_i < 1$. Se o padrão x_i for classificado incorretamente, a variável de folga correspondente terá valor $\xi_i > 1$. Ou seja, sempre que um padrão não estiver dentro de sua margem, a variável de folga ξ_i é positiva, conforme ilustrado na (FIGURA 6):

FIGURA 6 – SEPARAÇÃO LINEAR INEFICIENTE PARA DOIS CONJUNTOS DISTINTOS



FONTE: O autor (2016).

As variáveis de folga indicam a distância de um padrão em relação à margem correspondente a sua classe. Ou seja:

$$\begin{aligned} \underline{w}^t \underline{x} + b &\leq -1 - \xi_i, & x \in -1, \\ \underline{w}^t \underline{x} + b &\geq +1 - \xi_i, & x \in +1. \end{aligned} \quad (10)$$

Dessa maneira, é possível expressar a classificação correta através de formulação compacta:

$$y_i(\underline{w}^t \underline{x} + b) \geq +1 - \xi_i. \quad (11)$$

Este procedimento possibilita aceitar aqueles padrões que se situam fora da região de sua classe, impedindo que esses desviem o hiperplano separador $f(x)$.

Esse hiperplano é denominado hiperplano de margem flexível (CORTES; VAPNIK, 1995).

Assim, o problema primal utilizando as variáveis de folga é dado por:

$$\begin{aligned}
 \min \quad & \frac{1}{2} \underline{w}^t \underline{w} + C \sum_{i=1}^l \xi_i \\
 \text{s.a} \quad & \underline{w}^t \underline{x}_i + b \leq -1 - \xi_i, \quad x \in -1 \\
 & \underline{w}^t \underline{x}_i + b \geq +1 - \xi_i, \quad x \in +1 \\
 & \xi_i \geq 0 \quad i = 1, 2, \dots, l.
 \end{aligned} \tag{12}$$

Ou, de maneira compactada, dado por:

$$\begin{aligned}
 \min \quad & \frac{1}{2} \underline{w}^t \underline{w} + C \sum_{i=1}^l \xi_i \\
 \text{s.a} \quad & y_i (\underline{w}^t \underline{x} + b) \geq +1 - \xi_i \\
 & \xi_i \geq 0 \quad i = 1, 2, \dots, l.
 \end{aligned} \tag{13}$$

Onde \underline{w} , b são variáveis, $\underline{w} \in \mathbb{R}^n$, $b \in \mathbb{R}$ e C é um parâmetro que pondera os termos da minimização. O primeiro termo da função custo tem o objetivo de maximizar a margem, enquanto que o segundo termo, $C \sum_{i=1}^l \xi_i$, é utilizado para minimizar o valor das variáveis de folga ξ_i , reduzindo o número de pontos que ficam do lado incorreto. Assim, o parâmetro C atribui uma maior ou menor importância às variáveis de folga, possibilitando que o modelo do SVM seja menos sensível à presença de pontos “mal comportados” no conjunto de treinamento.

2.5.3 Modelo Dual

O problema de minimização representado no SVM normalmente é resolvido por meio de seu modelo dual. Segundo Morales e Simari (2010) apenas alguns trabalhos consideram o resolver diretamente o SVM pelo seu modelo primal.

O problema Lagrangeano dual é obtido acrescentando as restrições primais na função custo dual, por meio do uso de multiplicadores de Lagrange. Os

multiplicadores associados às restrições de desigualdade devem ser positivos e os associados à restrição de igualdade podem assumir quaisquer valores.

O problema Lagrangeano dual é conhecido como problema min-max, onde se deve minimizar a nova função de custo em relação aos parâmetros primal e maximizá-la em relação aos parâmetros dual, tendo única restrição, a expressão que garante que os Multiplicadores de Lagrange da restrição de desigualdade sejam positivos, ou seja, achar o ponto de sela do problema Lagrangeano dual.

Assim, dado o problema descrito pelo modelo primal, aplica-se a técnica do Lagrangeano, considerando o vetor $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_l)$ com os l multiplicadores de Lagrange e obtém-se então a função a ser maximizada dada por:

$$\begin{aligned} \text{Max } L(\underline{w}, b, \xi, \underline{\alpha}) &= W(\underline{w}, b, \xi) - \sum_{i=1}^l \alpha_i \left[y_i (\underline{w}^t \underline{x}_i + b) - 1 + \xi_i \right] \\ \text{S.a} \quad \alpha_i &\geq 0 \end{aligned} \quad (14)$$

Sendo α_i o Multiplicador de Lagrange associado a i -ésima restrição de desigualdade do modelo primal. Substituindo o valor de $W(\underline{w}, b, \xi)$, obtém-se:

$$\begin{aligned} \text{Max } L(\underline{w}, b, \xi, \underline{\alpha}) &= \frac{1}{2} \underline{w}^t \underline{w} + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i \left[y_i (\underline{w}^t \underline{x}_i + b) - 1 + \xi_i \right] \\ \text{S.a} \quad \alpha_i &\geq 0 \end{aligned} \quad (15)$$

Para minimizar o Lagrangeano dual em relação aos parâmetros primais, iguala-se a zero as derivadas parciais de primeira ordem em relação a cada um desses parâmetros, gerando as seguintes equivalências:

$$\frac{\partial L(\underline{w}, b, \xi, \underline{\alpha})}{\partial \underline{w}} = \underline{w} - \sum_{i=1}^l \alpha_i y_i \underline{x}_i = 0 \quad \Rightarrow \quad \underline{w} = \sum_{i=1}^l \alpha_i y_i \underline{x}_i, \quad (16)$$

$$\frac{\partial L(\underline{w}, b, \xi, \underline{\alpha})}{\partial b} = - \sum_{i=1}^l \alpha_i y_i = 0 \quad \Rightarrow \quad \sum_{i=1}^l \alpha_i y_i = 0, \quad (17)$$

$$\frac{\partial L(\underline{w}, b, \xi, \underline{\alpha})}{\partial \xi} = C - \sum_{i=1}^l \alpha_i = 0 \quad \Rightarrow \quad C = \sum_{i=1}^l \alpha_i. \quad (18)$$

Substituem-se as expressões obtidas no próprio Lagrangeano dual, obtendo-se um problema de maximização dual com restrição simples. Em virtude de obter apenas uma restrição e as variáveis limitadas o problema dual torna-se mais simples e mais fácil do que o problema primal.

$$\begin{aligned}
 \text{Max } L(\underline{w}, b, \xi, \underline{\alpha}) &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle \underline{x}_i, \underline{x}_j \rangle \\
 \text{S.a } \sum_{i=1}^l \alpha_i y_i &\geq 0 \\
 0 \leq \alpha_i &\leq C
 \end{aligned} \tag{19}$$

O valor α_i é o Multiplicador de Lagrange correspondente ao padrão i e C é o parâmetro que limita o valor dos Multiplicadores de Lagrange no ajuste do SVM.

De acordo com as condições de Karush-Kuhn-Tucker (KKT), as soluções ótimas $\underline{\alpha}^*$, (\underline{w}^*, b^*) devem satisfazer:

$$\alpha_i^* [y_i (\langle \underline{w}^t, \underline{x}_i \rangle + b) - 1 + \xi_i] = 0, \quad i = 1, \dots, l. \tag{20}$$

Isto implica que os α_i^* não nulos estão na margem funcional, chamados de vetores suportes e os demais são nulos (CRISTIANINI; SHAW-TAYLOR, 2000).

O Multiplicador de Lagrange associado a cada ponto torna-se uma variável dual, ou seja, o conjunto de treinamento é importante para a solução final. Os pontos que não são vetores suportes não afetam o resultado (CRISTIANINI; SHAW-TAYLOR, 2000).

Os valores nulos dos Multiplicadores de Lagrange não influenciam na obtenção dos pesos. Os padrões que têm o respectivo Multiplicador de Lagrange α não nulo são chamados de vetores suportes. Esses vetores são os únicos exemplos que contribuem para construção do hiperplano separador.

Uma característica do SVM é que ele apresenta esparsidade dos Multiplicadores de Lagrange ótimos, ou seja, após o treinamento do SVM vários Multiplicadores de Lagrange são nulos (CRISTIANINI; SHAW-TAYLOR, 2000).

Para classificar os padrões do conjunto de treinamento como vetores suporte verificam-se os valores dos respectivos Multiplicadores de Lagrange, sendo:

- i. Se $\alpha_i = 0$, $y_i \cdot f(\underline{x}_i) > 1$ então \underline{x}_i é considerado um vetor comum, que se situa do lado correto na região da sua classe e não influencia a construção do hiperplano separador ótimo;
- ii. Se $0 < \alpha_i < C$, $y_i \cdot f(\underline{x}_i) = 1$ então \underline{x}_i é um vetor suporte e situa-se sobre a margem da região da sua classe. É conhecido como vetor suporte *non-bound* (VS-NB);
- iii. E, se $\alpha_i = C$, $y_i \cdot f(\underline{x}_i) < 1$ então \underline{x}_i é um vetor suporte *bound* (VS-bound).

Assim, o padrão de treinamento pode se localizar entre a margem e o hiperplano separador, caso $0 < \xi_i < 1$; pode estar na própria superfície de separação, caso $\xi_i = 1$; ou ainda, pode estar na região de classe oposta a sua do outro lado da superfície de separação, caso $\xi_i > 1$ (CRISTIANINI; SHAW-TAYLOR, 2000).

O número de graus de liberdade do SVM depende da função *Kernel* escolhida. As funções *Kernel*, formulação e características são apresentadas no próximo capítulo, referente ao material e método. Algum conhecimento, *a priori*, pode ajudar no ajuste dos parâmetros da função *Kernel* escolhida.

Independente da classe de *Kernel* e dos dados é sempre possível encontrar os parâmetros do *Kernel* que torna os dados separáveis. Lembrando que forçar a separação de dados pode facilmente guiar a um particular *overfitting* quando os dados possuem ruídos (CRISTIANINI; SHAW-TAYLOR, 2000). O termo *overfitting* é utilizado quando o modelo estatístico se ajusta em demasiado ao conjunto de dados utilizados na fase de ajuste, ou seja, se adaptando perfeitamente aos dados de ajuste, podendo gerar erros além do normal para o conjunto de teste.

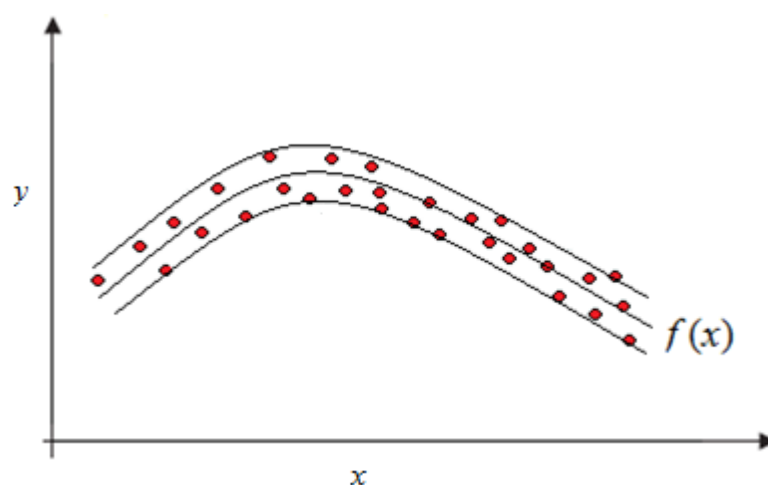
2.6 SUPPORT VECTOR REGRESSION (SVR)

Para problemas de regressão foi criada uma adaptação ou extensão, do método SVM. Tal extensão se dá pela inclusão de uma função de perda, denominada do inglês *loss function*.

A formulação do *Support Vector Regression* (SVR) utiliza os mesmos princípios do algoritmo de classificação (SVM), com exceção de alguns detalhes (PARRELLA, 2007).

Ao contrario do SVM, o SVR propõe determinar um hiperplano ótimo em que as amostras de treinamento estejam tão próximas quanto possível, não importando em qual dos lados da superfície os pontos se localizam e sim que a distância para a superfície seja a mínima possível. Porém, mesmo com propósitos opostos, ambos buscam estabelecer uma função com máxima capacidade de generalização (LIMA, 2004).

FIGURA 7 – HIPERPLANO ÓTIMO PARA A APLICAÇÃO DO SVR



FONTE: O autor (2015).

A (FIGURA 7) apresenta um hiperplano ótimo, onde todos os dados se situam entre as margens, e a linha central é a função $f(\underline{x})$ de regressão.

Sapankevych e Sankar (2009) destacam que a aplicação do SVR para a previsão de séries financeiras e de carga elétrica são as aplicações mais estudadas. Nas séries financeiras as análises são, mais especificamente, feitas para prever o índice do mercado de ações.

Novamente, considera-se um conjunto de treinamento descrito no qual $\underline{x}_i \subseteq \mathbb{R}^n$ é o vetor do espaço de entrada X . Porém, diferente do caso de classificação, como agora se trata de um problema de regressão, as saídas, y_i , desejadas podem assumir qualquer valor real, isto é, $y_i \subseteq \mathbb{R}$ para todo $i=1,2,\dots,l$.

O objetivo do SVR é encontrar uma função $f(\underline{x})$, com uma margem de erro caracterizada pelo intervalo $[y_i - \varepsilon, y_i + \varepsilon]$, onde, desvios são permitidos desde que não ultrapassem a margem especificada.

Assim, assumindo funções na forma $f(\underline{x}) = \langle \underline{w} \cdot \underline{x} \rangle + b$ tem-se, para o problema, as seguintes restrições:

$$\begin{aligned} y_i - \varepsilon &\leq \langle \underline{w} \cdot \underline{x}_i \rangle + b \Rightarrow y_i - \langle \underline{w} \cdot \underline{x}_i \rangle - b \leq \varepsilon, \\ y_i + \varepsilon &\geq \langle \underline{w} \cdot \underline{x}_i \rangle + b \Rightarrow \langle \underline{w} \cdot \underline{x}_i \rangle + b - y_i \leq \varepsilon. \end{aligned} \quad (21)$$

Visto que a função $f(\underline{x})$ deve satisfazer as restrições de erro $|f(\underline{x}_i) - y_i| \leq \varepsilon$, para todo $i=1,2,\dots,l$, modela-se o problema de otimização em sua forma primal como:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\underline{w}\|^2 \\ \text{s.a} \quad & y_i - \langle \underline{w} \cdot \underline{x}_i \rangle - b \leq \varepsilon \\ & \langle \underline{w} \cdot \underline{x}_i \rangle + b - y_i \leq \varepsilon \\ & \forall i = 1, \dots, l, \end{aligned} \quad (22)$$

onde $\underline{w} \in \mathbb{R}^n$ e $b \in \mathbb{R}$ são as incógnitas do problema.

O problema anterior supõe a existência de uma função $f(\underline{x})$ que aproxima todos os pares (\underline{x}_i, y_i) com uma precisão ε que deve ser especificada *a priori*. No entanto, nem sempre é possível garantir a viabilidade do problema anterior, visto que existem pontos que podem violar suas restrições. Assim, introduz-se uma função, conhecida como função de perda (*loss function*), que introduz variáveis de folgas não-negativas ξ_i, ξ_i^* , cuja finalidade é penalizar dados que se situem fora da margem $|f(\underline{x}_i) - y_i| \leq \varepsilon$. Considerando essa função é possível trabalhar com um

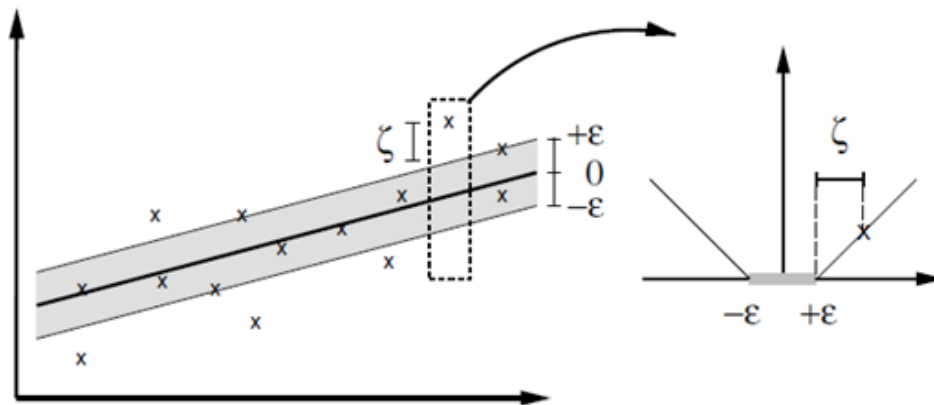
número limitado de erros, que em outras condições tornariam o problema de otimização inviável (DIAS, 2007).

A função de perda denominada ε – *Insensitive* é descrita por:

$$\begin{cases} |\xi|_{\varepsilon} = 0, & \text{se } |\xi| \leq \varepsilon \\ |\xi|_{\varepsilon} = |\xi| - \varepsilon, & \text{caso contrário} \end{cases} \quad (23)$$

A (FIGURA 8) mostra a representação gráfica da situação proposta pela função (23) na qual apenas os pontos localizados fora da região sombreada contribuem para o valor custo da função.

FIGURA 8 – FUNÇÃO PERDA ε – *INSENSITIVE*



FONTE: Smola e Schölkopf (2004).

As variáveis de folga ξ_i, ξ_i^* estão associadas aos dados que se situam fora das margens, tanto da margem inferior quanto da superior. Assim, pode-se reescrever o problema primal como sendo:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\underline{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s.a} \quad & y_i - \langle \underline{w} \cdot \underline{x}_i \rangle - b \leq \varepsilon + \xi_i \\ & \langle \underline{w} \cdot \underline{x}_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, \quad \forall i = 1, \dots, l, \end{aligned} \quad (24)$$

onde $\underline{w} \in \mathbb{R}^n$, $b \in \mathbb{R}$ são as incógnitas do problema e C é denominada constante de regularização, pois pondera os termos da função de minimização.

Assim como no caso do SVM, o método SVR também pode ser abordado pela sua formulação dual, que pode ser resolvido de maneira mais fácil, utilizando os multiplicadores de Lagrange, gerando o seguinte modelo de otimização:

$$\begin{aligned}
 \min \quad & -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle \underline{x}_i \cdot \underline{x}_j \rangle - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \\
 s.a \quad & \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\
 & \alpha_i, \alpha_i^* \in [0, C], \quad \forall i = 1, \dots, l,
 \end{aligned} \tag{25}$$

onde α_i, α_i^* são os multiplicadores de Lagrange.

Para o método SVR, as condições de complementaridade de Karush-Kuhn-Tucker (*KKT*) são descritas por:

$$\alpha_i (\varepsilon + \xi_i - y_i + \langle \underline{w} \cdot \underline{x}_i \rangle + b) = 0, \quad \forall i = 1, \dots, l, \tag{26}$$

$$\alpha_i^* (\varepsilon + \xi_i^* - y_i + \langle \underline{w} \cdot \underline{x}_i \rangle + b) = 0, \quad \forall i = 1, \dots, l, \tag{27}$$

$$(C - \alpha_i) \xi_i = 0, \quad \forall i = 1, \dots, l, \tag{28}$$

$$(C - \alpha_i^*) \xi_i^* = 0, \quad \forall i = 1, \dots, l, \tag{29}$$

$$\alpha_i \alpha_i^* = 0, \quad \forall i = 1, \dots, l, \tag{30}$$

$$\xi_i \xi_i^* = 0, \quad \forall i = 1, \dots, l. \tag{31}$$

Analisando as condições de *KKT*, novamente extraem-se informações importantes quanto à solução do modelo. Pela equação (30), observa-se que não é possível a existência de um conjunto de variáveis duais α_i e α_i^* , em que ambos os valores sejam não nulos.

Das relações (26) e (27) obtém-se que apenas os pontos de treinamento em que $|f(\underline{x}_i) - y_i| \geq \varepsilon$ estão associados aos multiplicadores de Lagrange não nulos. Tais pontos estão localizados sobre as margens $+\varepsilon$ e $-\varepsilon$ ou fora da região

delimitada. Estes dados são os únicos a serem utilizados no cálculo do vetor dos pesos \underline{w} e por isso são chamados de vetores suporte.

Pelas condições (28) e (29) os vetores suporte cujo $\alpha_i = C$ ou $\alpha_i^* = C$, têm respectivamente variáveis de folga ξ_i ou ξ_i^* não nulas. Caracterizando pontos localizados fora das margens e correspondem a erros do modelo.

A expressão do vetor de pesos é calculado pela expressão (GUNN, 1998):

$$\underline{w}^* = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \underline{x}_i \quad (32)$$

e, a função de decisão é dada por:

$$f(\underline{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle \underline{x}_i \cdot \underline{x} \rangle + b. \quad (33)$$

onde o *bias* pode ser determinado por meio da relação:

$$b^* = -\frac{1}{2} \langle \underline{w}^*, (\underline{x}_r + \underline{x}_s) \rangle \quad (34)$$

em que \underline{x}_r e \underline{x}_s são quaisquer vetores suporte de cada classe, satisfazendo $\alpha_r, \alpha_s > 0$ e $y_r = -1$, $y_s = +1$.

2.7 LEAST SQUARES SUPPORT VECTOR MACHINE (LS-SVM)

Proposto por Suykens e Vandewalle (1999), LS-SVM é uma variante do método SVM clássico. *Least Squares Support Vector Machine* (Máquina de Vetores Suporte a Mínimos Quadrados) mantém as mesmas características básicas e a mesma qualidade na solução encontrada que a sua predecessora (SUYKENS, *et al.* 2002). Ao contrário do SVM, o LS-SVM considera restrições de igualdade no lugar das desigualdades, com isso, resulta um algoritmo que reduz os problemas ao se aplicar a um conjunto extenso de dados (SANTOS, 2013).

Uma das maiores diferenças é que ao contrário do SVM que utiliza a programação quadrática para calcular seus vetores suporte, que demanda um

grande tempo computacional e a obtenção da solução ótima que possui complexidade considerável, o LS-SVM usa um sistema de equações lineares e a função de custo por mínimos quadrados (BORIN, 2007; SHAH, 2005).

O LS-SVM pode ser aplicado tanto a problemas de classificação binária ou multi-classes quanto para problemas de regressão, assim como o método SVM.

Portanto, como para o método SVM tradicional, dado um conjunto de dados de entrada $S = \{(\underline{x}_1, y_1), (\underline{x}_2, y_2), \dots, (\underline{x}_l, y_l)\} \subseteq \{X \times Y\}^l$ para o treinamento, ou ajuste do modelo, o objetivo é estimar os parâmetros \underline{w} e b do modelo:

$$y = \underline{w}^t \underline{x} + b, \quad (35)$$

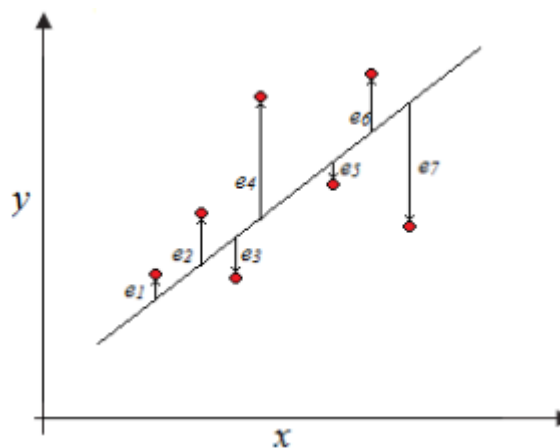
com $\underline{x} \in \mathbb{R}^n$, $y \in \mathbb{R}$, $i = 1, 2, \dots, n$, \underline{w} são os pesos e b o *bias*.

Logo, formula-se o problema primal:

$$\begin{aligned} \min \quad & \frac{1}{2} \underline{w}^t \underline{w} + \frac{C}{2} \sum_{i=1}^n e_i^2 \\ \text{s.a} \quad & y_i = \underline{w}^t \underline{x}_i + b + e_i \\ & i = 1, 2, \dots, n, \end{aligned} \quad (36)$$

com \underline{w} sendo o vetor de pesos, C o parâmetro que penaliza erros altos e é otimizado pelo usuário e e_i são os erros mínimos em relação à reta de regressão, conforme se pode verificar no exemplo da (FIGURA 9) dada a seguir.

FIGURA 9 – EXEMPLO DE ERROS MÍNIMOS EM RELAÇÃO À FUNÇÃO DE REGRESSÃO



FONTE: O autor (2016).

Aplicando o Lagrangeano tem-se o seguinte modelo:

$$L(\underline{w}, b, e; \underline{\alpha}) = \frac{1}{2} \underline{w}^t \underline{w} + \frac{C}{2} \sum_{i=1}^n e_i^2 - \sum_{i=1}^n \alpha_i (\underline{w}^t \underline{x}_i + b + e_i - y_i), \quad (37)$$

onde $\alpha_i \in \mathbb{R}$ são os multiplicadores de Lagrange e as condições de otimalidade são dadas por:

$$\frac{\partial L}{\partial \underline{w}} = 0 \quad \Rightarrow \quad \underline{w} = \sum_{i=1}^n \alpha_i \underline{x}_i, \quad (38)$$

$$\frac{\partial L}{\partial b} = 0 \quad \Rightarrow \quad \sum_{i=1}^n \alpha_i = 0, \quad (39)$$

$$\frac{\partial L}{\partial e_i} = 0 \quad \Rightarrow \quad \alpha_i = C e_i, \quad (40)$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \quad \Rightarrow \quad y_i = \underline{w}^t \underline{x}_i + b + e_i. \quad (41)$$

Substituindo \underline{w} e e no Lagrangiano obtém-se o seguinte sistema de equações lineares *KKT* (SHAH, 2005):

$$\begin{bmatrix} 0 & \underline{1}_n^t \\ \underline{1}_n & (\underline{x}_i \cdot \underline{x}_j) + I_n / C \end{bmatrix} \cdot \begin{bmatrix} b \\ \underline{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \underline{Y} \end{bmatrix}, \quad (42)$$

com $\underline{Y} = (y_1, y_2, \dots, y_n)^t$, $\underline{1}_n = (1, 1, \dots, 1)^t$ e $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)^t$.

Assim com o SVM e o SVR pode-se utilizar a técnica da função núcleo, *Kernel*, para tratar de padrões que não são linearmente separáveis no espaço de entrada. Tal abordagem e modificações serão tratadas adiante, após a definição de função *Kernel*.

2.8 FUNÇÕES *KERNEL*

Em alguns problemas, podem aparecer dados de padrões diferentes muito próximos, uns aos outros, impedindo totalmente ou gerando um hiperplano separador sujeito a erros devido à proximidade dos padrões de classes distintas. Assim, a separação dos padrões e a sua classificação fica comprometida, não sendo o modelo capaz de classificá-los linearmente.

Tal problema inviabiliza a separação linear dos dados, forçando a introdução de um artifício com que se possa contornar essa limitação. Assim, uma saída é, utilizando o artifício das funções *Kernel* fazer um mapeamento dos dados não linearmente separáveis em um espaço de dimensão maior, onde os dados tornam-se linearmente separáveis (CARVALHO, 2005a).

A partir dos dados originais no espaço de entrada, a utilização da função *Kernel* consiste em transformar tais dados de dimensão n , em dados com uma dimensão p maior que a original.

Quando se faz uso de uma função $\phi: IR^n \rightarrow IR^p$, com $p > n$, os dados estão sendo mapeados no espaço *feature* (espaço das características). O espaço *feature* consiste em um espaço no qual cada vetor de entrada é mapeado, por meio da função $\phi(\underline{x}_i) \in IR^p$, em que p pode possuir qualquer valor maior do que n , dimensão do espaço de entrada.

Considere um mapeamento que permite construir uma função polinomial de decisão no espaço de entrada. Para se determinar um polinômio de grau dois, pode-se criar um espaço de características, onde \underline{x} é um vetor no espaço de entrada com n coordenadas, podendo ser escrito como $\underline{x} = (x_1, x_2, \dots, x_n)$ e o espaço *feature* teria suas coordenadas da seguinte maneira (VAPNIK, 1999):

$$\begin{aligned}
 z_1 = x_1, \dots, z_n = x_n & \quad (n \text{ coordenadas}); \\
 z_{n+1} = (x_1)^2, \dots, z_{2n} = (x_n)^2 & \quad (n \text{ coordenadas}); \\
 z_{2n+1} = x_1 x_2, \dots, z_N = x_n x_{n-1} & \quad \left(\frac{n(n-1)}{2} \text{ coordenadas} \right).
 \end{aligned} \tag{43}$$

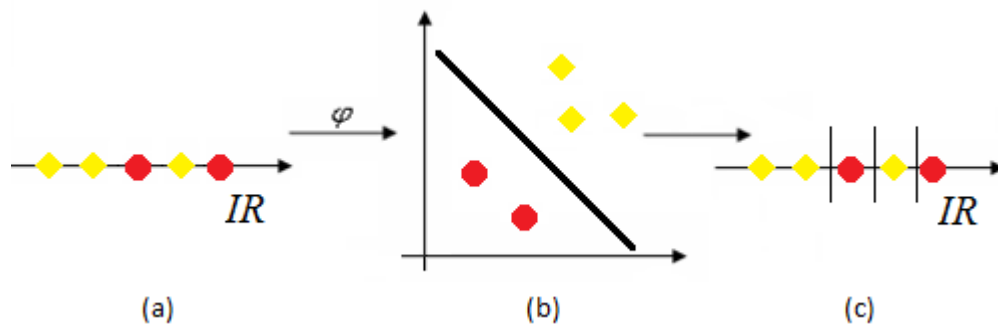
Portanto, um espaço de dimensão n seria levado a um espaço *feature* de dimensão $N = \frac{n(n+3)}{2}$. Se o problema tem dimensão pequena, é possível construir o espaço *feature*. No entanto, se o problema tiver dimensão maior, por exemplo, \mathbb{R}^{200} , um polinomial de grau cinco geraria bilhões de coordenadas tornando-se difícil de ser tratado computacionalmente (VAPNIK, 1999).

Na função objetivo do problema dual do SVM existe um produto interno entre os pontos, o qual pode ser substituído por uma única função denominada função Kernel. O Kernel é uma função com a seguinte característica:

$$K(x_i, x_j) = K(x_j, x_i) = \varphi(x_i) \cdot \varphi(x_j), \quad (44)$$

para x_i e x_j pertencentes ao espaço de entrada, onde φ é um mapeamento do espaço de entrada no espaço das *features* (FIGURA 10).

FIGURA 10 – MAPEAMENTO REALIZADO PELA FUNÇÃO KERNEL



FONTE: O autor (2016).

Na (FIGURA 10), os dados que estão no espaço de entrada não são separáveis por um hiperplano (FIGURA 10a). A função φ faz uma transformação nos dados de maneira que, no espaço *feature*, os dados tornam-se linearmente separáveis (FIGURA 10b) e retornando ao espaço de entrada, já linearmente separados os dados (FIGUR 10c).

O hiperplano de separação é criado no espaço *feature*, no qual os dados tendem a ser linearmente separáveis (FIGURA 10b). Porém, a separação é rebatida para o espaço de entrada, utilizando a superfície de decisão $f(\underline{x})=0$ que ocorre no

próprio espaço de entrada, de forma a separar os dados em classes diferentes (FIGUR 10c) (CARVALHO, 2005b).

Caso um hiperplano consiga generalizar a separação e teoricamente possa ser encontrado, resta o problema técnico de como tratar a alta dimensão do espaço *feature*. No entanto, para construção do hiperplano de separação ótimo no espaço *feature*, não é necessário considerá-lo de forma explícita.

Considerando a propriedade geral de produto interno no espaço de Hilbert, supõe-se um mapeamento ϕ do vetor $\underline{x} \in \mathbb{R}^n$, $\phi: \mathbb{R}^n \rightarrow H$, dentro do espaço de Hilbert com as seguintes coordenadas $\phi_1(\underline{x}), \phi_2(\underline{x}), \dots, \phi_n(\underline{x}), \dots$. Espaço de Hilbert H é um espaço vetorial de dimensão finita ou infinita. De acordo com a Teoria de Hilbert-Schmidt, o produto interno, no espaço de Hilbert, tem representação equivalente a:

$$\langle \underline{\phi}_1 \cdot \underline{\phi}_2 \rangle = \sum_{p=1}^{\infty} \lambda_p \phi_p(\underline{x}_i) \phi_p(\underline{x}_j) \Leftrightarrow K(\underline{x}_i, \underline{x}_j). \quad (45)$$

Uma função é dita ser uma função *Kernel* se ela satisfaz as condições estabelecidas pelo Teorema de Mercer (CARVALHO, 2005a).

Teorema de Mercer: Uma função é dita ser uma função *Kernel*, se a matriz K é positiva definida, onde K é obtida por:

$$K = K_{ij} = K(\underline{x}_i, \underline{x}_j). \quad (46)$$

Uma matriz é positiva definida, se seus autovalores são maiores que zero. As funções *Kernel* que satisfazem as condições do Teorema de Mercer são chamadas de *Kernels* de Mercer (SMOLA *et al.*, 1999).

O produto interno no espaço de Hilbert que conduz a construção do SVM implica que para qualquer função *Kernel* satisfazendo a condição de Mercer existe um espaço *feature* onde as funções geram um produto interno (VAPNIK, 1999).

Assim, tem-se que, uma função será *Kernel* se ela for simétrica,

$$K(\underline{x}, \underline{z}) = \langle \underline{\phi}(\underline{x}) \cdot \underline{\phi}(\underline{z}) \rangle = \langle \underline{\phi}(\underline{z}) \cdot \underline{\phi}(\underline{x}) \rangle = K(\underline{z}, \underline{x}) \quad (47)$$

e satisfazer a desigualdade de Cauchy-Schwarz,

$$K(\underline{x}, \underline{z})^2 = \langle \underline{\phi}(\underline{x}) \cdot \underline{\phi}(\underline{z}) \rangle^2 \leq \| \underline{\phi}(\underline{x}) \|^2 \cdot \| \underline{\phi}(\underline{z}) \|^2 = \langle \underline{\phi}(\underline{x}) \cdot \underline{\phi}(\underline{z}) \rangle \langle \underline{\phi}(\underline{x}) \cdot \underline{\phi}(\underline{x}) \rangle \leq K(\underline{x}, \underline{x}) K(\underline{z}, \underline{z}) . \quad (48)$$

Não se pode dizer que uma função *Kernel* é melhor que outra. O desempenho de uma função vai depender do tipo de banco de dados utilizado na aplicação. Sendo assim, é possível testar diversos tipos de *Kernel* a fim de encontrar aquela que melhor se adapte ao problema estudado (NGUYEN *et al.*, 2006).

A seguir são apresentadas algumas funções *Kernel* mais utilizadas:

i. Produto interno (*Kernel* linear padrão):

$$K(\underline{x}_i, \underline{x}_j) = \underline{x}_i^t \cdot \underline{x}_j . \quad (49)$$

ii. Polinomial Homogêneo:

$$K(\underline{x}_i, \underline{x}_j) = \left(\underline{x}_i^t \cdot \underline{x}_j \right)^p , \quad (50)$$

onde p é o grau do polinômio.

iii. Polinomial não homogêneo:

$$K(\underline{x}_i, \underline{x}_j) = \left(\underline{x}_i^t \cdot \underline{x}_j + k \right)^p , \quad (51)$$

onde

p é o grau do polinômio;

k é uma constante.

iv. Sigmoidal:

$$K(\underline{x}_i, \underline{x}_j) = \tanh \left(\kappa \cdot \underline{x}_i^t \cdot \underline{x}_j + k \right) , \quad (52)$$

onde

κ é um coeficiente;

k é uma constante negativa.

Segundo Hsu *et al.* (2003) essa função *Kernel* não é válida para alguns parâmetros, já que não se trata de um produto interno.

v. Gaussiano (RBF):

$$K(\underline{x}_i, \underline{x}_j) = e^{-\frac{\|\underline{x}_i - \underline{x}_j\|^2}{2\sigma^2}}, \quad (53)$$

onde σ é um parâmetro.

O *Kernel* Gaussiano também é conhecido como RBF (*Radial Basis Function* – Funções de Base Radial). Hsu *et al.* (2003) afirmam que o *Kernel* RBF é o mais adequado, pois possui valores menores para os *Kernels*, ou seja, $0 \leq K(\underline{x}_i, \underline{x}_j) \leq 1$, enquanto os demais podem divergir, como, por exemplo, o *Kernel* polinomial, que pode variar de zero a infinito, caso o valor do grau seja elevado.

Qualquer outra função que atenda à condição de Mercer pode ser utilizada como função *Kernel*, assim, existe uma infinidade de funções *Kernel* a serem propostas.

2.8.1 *Kernel* Aplicado ao Modelo LS-SVM

Com a introdução das funções *Kernel*, o tratamento do modelo muda apenas na substituição do produto interno pela função *Kernel* escolhida. Assim, o modelo primal Lagrangeano continua o mesmo:

$$L(\underline{w}, b, e; \underline{\alpha}) = \frac{1}{2} \underline{w}^t \underline{w} + \frac{C}{2} \sum_{i=1}^n e_i^2 - \sum_{i=1}^n \alpha_i (\underline{w}^t \underline{\phi}(x_i) + b + e_i - y_i). \quad (54)$$

E o sistema linear de equações

$$\begin{bmatrix} 0 & \underline{1}_n^t \\ \underline{1}_n & \Omega + I_n/C \end{bmatrix} \begin{bmatrix} b \\ \underline{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \underline{Y} \end{bmatrix} \quad (55)$$

com $\underline{Y} = (y_1, y_2, \dots, y_n)^t$, $\underline{1}_n = (1, 1, \dots, 1)^t$, $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)^t$ e Ω é uma matriz dada pelos elementos $\Omega = \Omega_{ij} = \phi(\underline{x}_i) \phi(\underline{x}_j) = K(\underline{x}_i, \underline{x}_j)$ com $i, j = 1, \dots, n$. E o modelo dual LS-SVM fica

$$y = f(\underline{x}) = \sum_{i=1}^n \alpha_i K(\underline{x}, \underline{x}_i) + b. \quad (56)$$

2.9 MÉTODO *BOOTSTRAP*

O *Bootstrap* é uma técnica estatística computacionalmente intensiva não paramétrica, que permite a avaliação da variabilidade de estatísticas, com base nos dados de uma amostra existente inicial. É indicado para problemas onde os procedimentos estatísticos padrões não existam ou que sejam de difícil aplicação, e é vantajoso se usado em problemas com amostra finita ou, ainda, com amostra grande desde que forneça resultado semelhante ou superior ao método assintótico usual (CHAVES NETO, 1991).

O *Bootstrap* surgiu quando Efron em 1979 (*apud* CHAVES NETO, 1991) estudava o problema da estimação da distribuição amostral de uma estatística $T_n(\underline{x}, F)$ com base nos dados de uma amostra de tamanho n , $\underline{x} = [x_1, x_2, \dots, x_n]$, de uma distribuição de probabilidade desconhecida F , com $X_i \sim i.i.d.F$.

O procedimento *Bootstrap* é desenvolvido a partir de uma amostra inicial $\underline{x} = [x_1, x_2, \dots, x_n]$ da qual se toma B novas amostras ou reamostras conhecidas como pseudo-dados, de mesmo tamanho com reposição. Então, calcula-se a estatística de interesse para cada reamostra *Bootstrap* obtendo-se B valores *Bootstrap* da estatística.

Assim o conjunto com cardinalidade B de valores *Bootstrap* da estatística corresponde a uma estimativa da verdadeira distribuição amostral da estatística em questão. Essa operação pode ser representada pelo seguinte esquema:

Seja o parâmetro θ e seu estimador a estatística $T_n(\underline{x}, F)$, onde $\underline{x} = [x_1, x_2, \dots, x_n]$ é a amostra aleatória disponível da variável aleatória X_i com função distribuição desconhecida F , com $X_i \sim i.i.d.F$.

Então:

1º) o estimador não-paramétrico de máxima verossimilhança $\hat{F}_n(x)$ de F é:

$$\hat{F}_n(x) = \left[\sum_{i=1}^n I(X_i \leq x) / n \right] \quad (57)$$

com $I(X_i \leq x)$ sendo função indicadora. Assim essa distribuição empírica é construída colocando-se massa probabilística $1/n$ em cada ponto amostral;

2º) de \hat{F}_n toma-se B (muito grande) amostras *bootstrap* de mesmo tamanho n ,

$$\begin{aligned}
 \underline{x}_1^* &= [x_{11}^*, x_{12}^*, \dots, x_{1n}^*] & X_{1i}^* &\sim iid \hat{F}_n \\
 \underline{x}_2^* &= [x_{21}^*, x_{22}^*, \dots, x_{2n}^*] & X_{2i}^* &\sim iid \hat{F}_n \\
 &\dots\dots\dots & & \dots\dots\dots \\
 \underline{x}_B^* &= [x_{B1}^*, x_{B2}^*, \dots, x_{Bn}^*] & X_{Bi}^* &\sim iid \hat{F}_n
 \end{aligned} \tag{58}$$

3º) calcula-se as B estatísticas *Bootstrap* $T_n^*(\underline{x}_l^*)$ com $l=1,2,\dots,B$, correspondentes às B amostras *bootstrap* e forma-se o conjunto:

$$\{T_n^*(\underline{x}_l^*); l=1,2,\dots,B\}, \tag{59}$$

que é uma simulação da verdadeira distribuição amostral da estatística $T_n(\underline{x}, F)$.

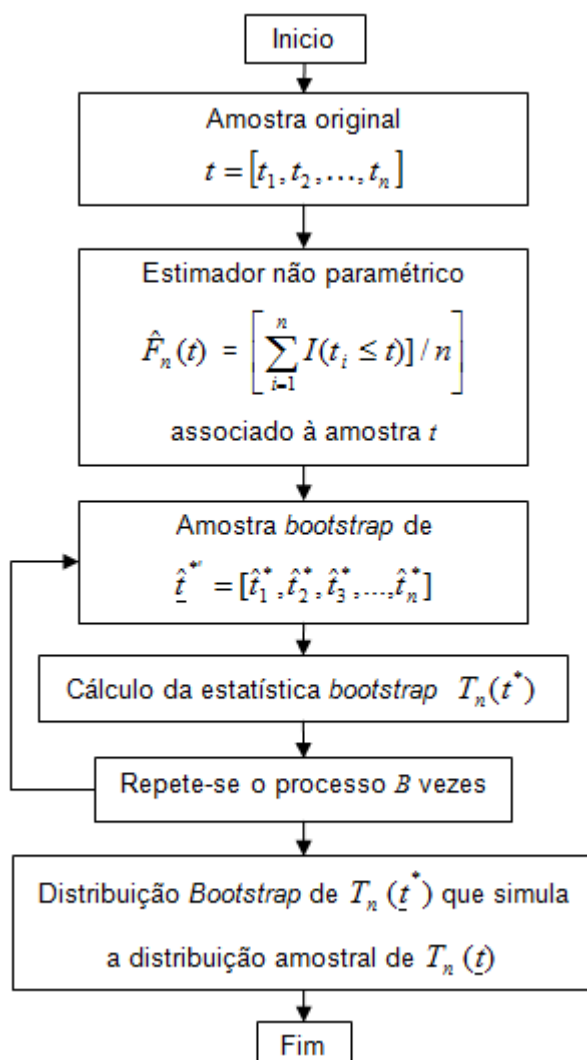
Assim, com o conjunto $\{T_n^*(\underline{x}_l^*); l=1,2,\dots,B\}$ disponível pode-se obter a uma medida da variabilidade de $T_n(\underline{x}, F)$, tal como o seu erro-padrão *Bootstrap*. A estimativa do vício da estatística $T_n(\underline{x}, F)$ é dada por:

$$b[\theta, T_n(\underline{x}, F)] = T_n(\underline{x}, F) - T_n^*, \tag{60}$$

onde T_n^* corresponde à estimativa *bootstrap* do parâmetro e que é dada por:

$$T_n^* = \frac{[\sum_{l=1}^B T_n(\underline{x}_l^*)]}{B}. \tag{61}$$

Na prática, constrói-se a distribuição *Bootstrap*, da estatística $T_n(\underline{x}, F)$, pelo método de Monte-Carlo com um número de replicações, B , suficientemente grande. Um indicador do tamanho adequado de B , independente do custo computacional, é a qualidade da convergência da estimativa *Bootstrap* do parâmetro para a estimativa natural do parâmetro ($T_n^* \xrightarrow{B \rightarrow \infty} T_n(\underline{x}, F)$). O fluxograma da (FIGURA 11) apresenta o algoritmo de construção da distribuição *bootstrap*.

FIGURA 11 – ALGORITMO DA DISTRIBUIÇÃO BOOTSTRAP DA ESTATÍSTICA $T_n(\underline{x}, F)$ 

FONTE: Chaves Neto (1991).

O procedimento descrito se aplica ao método *Bootstrap* Não-Paramétrico. Ao se optar pela utilização do método *Bootstrap* Paramétrico, procede-se da mesma forma, com a diferença de que cada amostra *bootstrap* é obtida da distribuição paramétrica que originou os dados, ao invés de reamostrar as observações disponíveis.

2.10 FUNÇÃO DISCRIMINANTE LINEAR DE FISHER

A função discriminante linear de Fisher é uma combinação linear de características originais a qual se caracteriza por produzir separação máxima entre duas populações.

Considerando que $\underline{\mu}_i$ e Σ são parâmetros conhecidos e correspondem respectivamente aos vetores de médias e a matriz de covariâncias comum das populações π_i . Demonstra-se que a função linear do vetor aleatório \underline{X} que produz separação máxima entre duas populações é dada por:

$$D(\underline{X}) = \underline{L}' \underline{X} = [\underline{\mu}_1 - \underline{\mu}_2]' \Sigma^{-1} \underline{X}, \quad (62)$$

onde:

$\underline{X} = [X_1, \dots, X_p]$ e $\underline{\pi} = [\pi_1, \pi_2]$;

\underline{L} = vetor discriminante;

\underline{X} = vetor aleatório de características das populações;

$\underline{\mu}$ = vetor de médias p-variado;

Σ = matriz comum de covariâncias das populações π_1 e π_2 .

O valor da função discriminante de Fisher para uma dada observação \underline{x}_0 é dada por:

$$D(x_0) = [\underline{\mu}_1 - \underline{\mu}_2]' \Sigma^{-1} \underline{x}_0. \quad (63)$$

O ponto médio entre as duas médias populacionais univariadas μ_1 e μ_2 é:

$$m = \frac{1}{2} [\underline{\mu}_1 - \underline{\mu}_2]' \Sigma^{-1} [\underline{\mu}_1 + \underline{\mu}_2], \quad (64)$$

$$m = \frac{1}{2} [D(\underline{\mu}_1) + D(\underline{\mu}_2)]. \quad (65)$$

A regra de classificação baseada na função discriminante de Fisher é dada por:

$$\underline{x}_0 \in \pi_1 \text{ se } D(\underline{x}_0) = [\underline{\mu}_1 - \underline{\mu}_2]' \Sigma^{-1} \underline{x}_0 \geq m, \quad (66)$$

$$\underline{x}_0 \in \pi_2 \text{ se } D(\underline{x}_0) = [\underline{\mu}_1 - \underline{\mu}_2]' \Sigma^{-1} \underline{x}_0 < m. \quad (67)$$

Considerando-se que as populações π_1 e π_2 têm a mesma matriz de covariâncias Σ pode-se então, estimar a matriz comum de covariâncias S_c :

$$S_c = \left[\frac{n_1 - 1}{(n_1 - 1) + (n_2 - 1)} \right] S_1 + \left[\frac{n_2 - 1}{(n_1 - 1) + (n_2 - 1)} \right] S_2, \quad (68)$$

onde,

S_c = estimativa da matriz comum de covariâncias;

n_1 = número de observações da população π_1 ;

n_2 = número de observações da população π_2 ;

\underline{X} = vetor aleatório de características das populações;

S_1 = estimativa da matriz de covariâncias da população π_1 ;

S_2 = estimativa da matriz de covariâncias da população π_2 .

A função discriminante linear de Fisher é obtida substituindo-se os parâmetros $\underline{\mu}_1$ e $\underline{\mu}_2$ e Σ pelas respectivas quantidades amostrais $\bar{\underline{x}}_1$, $\bar{\underline{x}}_2$ e S_c :

$$D(\underline{x}) = \hat{L}' \underline{x} = [\bar{\underline{x}}_1 - \bar{\underline{x}}_2]' S_c^{-1} \underline{x}, \quad (69)$$

onde,

$D(\underline{x})$ = função discriminante de linear amostral de Fisher;

\hat{L}' = estimativa do vetor discriminante;

$\bar{\underline{x}}_1$ = média amostral da população π_1 ;

$\bar{\underline{x}}_2$ = média amostral da população π_2 .

3 REVISÃO DE LITERATURA

São apresentados neste capítulo artigos com aplicações dos métodos SVM, SVR e LS-SVM, encontrados na literatura. Destacam-se aplicações em áreas tais como a biometria e segurança, nas instituições de crédito, medicina e biologia e na indústria.

3.1 BIOMETRIA E SEGURANÇA

3.1.1 Classificação de Faces e Assinaturas

Li *et al.* (2000) apresentaram os resultados de um trabalho onde desenvolveram uma metodologia de reconhecimento de faces utilizando o *Support Vector Regression*. O método foi utilizado para calcular a posição do rosto na imagem e, assim que o encontra usa o SVM multi-classes para fazer o reconhecimento facial.

Pela característica da face humana em ser quase simétrica, em relação ao nariz, as imagens do lado direito da face são invertidas e tornam-se imagens do lado esquerdo da face. Isso ajudou a simplificar a complexidade da modelagem, sem perda das características gerais. Os resultados chegaram a 95% na taxa de detecção da posição do rosto e a 90% de precisão do reconhecimento.

Özgündüz, Sentürk e Karsligil (2005) apresentaram um sistema *off-line* de reconhecimento e verificação de assinaturas. No pré-processamento, a assinatura foi escaneada. Em seguida, eliminou-se a imagem do seu fundo, reduzindo-se os ruídos, ajustou-se a um tamanho padrão e eliminou-se a diferença de espessura provocada pela caneta utilizada e pela pressão aplicada, passando para a etapa de extração das características da assinatura.

As principais características observadas e utilizadas para montar os padrões são: a área de assinatura, densidade da assinatura, direção das linhas, frequência horizontal e vertical, número de cantos. Tal procedimento trata-se de um problema multiclases, assim, empregou-se o método um contra todos. O conjunto de dados disponíveis era composto de 1320 assinaturas coletadas de um grupo de 70

peçoas, sendo verdadeiras e falsas, e a taxa de classificação ficou em cerca de 95%.

3.2 FINANÇAS

3.2.1 Instituições de Crédito

Lai *et al.* (2006) utilizaram o LS-SVM, modificação do SVM, para executar uma análise na avaliação do risco de crédito, comparando com Regressão Linear, Regressão Logística, Redes Neurais Artificiais e SVM.

O problema era classificar os credores em adimplentes e inadimplentes. Foi utilizado um conjunto de dados com 1225 clientes, sendo 323 inadimplentes. Ao todo foram observadas 14 variáveis, como idade, número de filhos, dependentes, rendimento mensal, valor do imóvel, etc.

Os autores obtiveram resultados melhores quando utilizou-se o método LS-SVM com o *Kernel* Gaussiano (RBF).

3.2.2 Previsões de Séries Temporais Financeiras

Trafalis e Ince (2000) apresentaram um estudo comparativo entre o SVR e as redes neurais *Back-propagation* e de Base Radial aplicado à previsão de preços de ações. Nas fases de treinamento e de teste dos modelos foram utilizados respectivamente os conjuntos de dados formados pelos preços diários das ações da IBM (*International Business Machines*), Yahoo e AOL (*America Online*) negociadas nos períodos de 07/05 à 19/10 do ano de 1998. Os resultados preliminares da pesquisa foram considerados promissores.

Os desempenhos de ambos os modelos variaram conforme a seleção de seus parâmetros, de forma que, em determinadas situações, as previsões realizadas pelo SVR mostraram-se superiores às realizadas pelas redes neurais e, em outras, mostraram-se inferiores.

3.3 MEDICINA

Yu *et al.* (2006) aplicaram o SVM para classificar os sinais dos exames de eletrocardiograma (ECG) e fotopletismografia (FPG) como sendo adequados ou inadequados. Foram avaliadas 158 amostras de pacientes com trauma, selecionadas aleatoriamente, coletadas durante o transporte de helicóptero. Cada amostra foi composta de 7 segundos do exame ECG e segmentos do exame FPG e a taxa de referência da confiabilidade do coração.

Foram comparados os resultados do algoritmo de análise quantitativa contra análises qualitativas, realizada por peritos humanos, e verificou que em 92% dos casos o SVM apresentou concordância com a avaliação feita pelos médicos, apresentando um resultado satisfatório.

3.4 INDÚSTRIA

3.4.1 Previsão de Demanda de Energia Elétrica

Ruas *et al.* (2007) aplicaram a metodologia do SVR e de Redes Neurais Artificiais para calcular previsões de demanda em pontos de conexão da rede de energia em curto prazo, cerca de 15 minutos, no estado do Paraná, entre 2004 e 2006. Ambas as metodologias obtiveram desempenho satisfatório, comparável com os publicados na literatura. Porém, para as redes neurais, existiu uma desvantagem: o uso da definição automática da estrutura da rede elétrica emprega um grande número de testes e muita experiência por parte do projetista.

Assim, a metodologia do SVR mostrou-se viável, uma vez que exigiu apenas o ajuste no parâmetro C .

3.4.2 Quantificar os Adulterantes no Leite em Pó

Ferrão *et al.* (2007) apresentaram uma aplicação do LS-SVM para analisar as quantidades de adulterantes no composto do leite em pó, uma vez que a exigência na qualidade passou a patamares elevados, após o crescimento das exportações.

Foram apresentadas amostras de leite em pó contaminadas com pó de amido, sacarose e com soro de leite.

Os resultados da aplicação do LS-SVM foram comparados com os resultados da técnica de regressão por mínimos quadrados parciais, onde o LS-SVM se sobressaiu tanto nos resultados como na capacidade de generalização.

3.5 PREVISÃO DA EVAPORAÇÃO DE ÁGUA EM RESERVATÓRIOS

Samui (2011) aborda em seu artigo uma aplicação do método SVR utilizando a modelagem LS-SVR, para a predição de perdas por evaporação, em reservatórios de água abertos. A entrada do modelo LS-SVR foi formada por valores da temperatura média do ar (°C), velocidade média do vento (m/seg), horas de sol por dia e umidade relativa do ar (%).

A base de dados foi obtida do trabalho de Deswal e Pal (2008), que aplicou uma rede neural artificial (RNA) aos dados para se obter modelos de previsão da evaporação de água. Tanto os valores obtidos em Samui (2011), como em Deswal e Pal (2008), foram comparados com os resultados obtidos pela metodologia desenvolvida nesta tese, mais adiante.

4 MATERIAL E MÉTODO

Neste capítulo são apresentados formalmente a metodologia aplicada no trabalho e o material, ou seja, as séries temporais estudadas e às quais foi aplicado o desenvolvimento das metodologias a serem apresentadas nas seções a seguir.

4.1 CONFIGURAÇÃO DOS CONJUNTOS DE DADOS

Os *softwares* utilizados na elaboração das análises estão descritos a seguir:

- i) Todas as abordagens dos problemas foram implementadas em linguagem Fortran, sendo utilizado como auxílio aos cálculos o *software* MATLAB R2012b;
- ii) Para a elaboração do programa computacional em linguagem FORTRAN *Free Form* (Fortran 95) foram utilizados os compiladores *Microsoft Developer Studio* e *Microsoft Visual Studio* 2008;
- iii) Para a elaboração dos gráficos: MATLAB R2012b.

Como o SVM é uma técnica de treinamento supervisionado, devem-se estabelecer alguns critérios antes do início do treinamento para certo conjunto de dados. Geralmente, quando se tem um conjunto grande de dados, divide-se tal conjunto em três subconjuntos menores (CAMILO; SILVA, 2009):

1. Conjunto de Treinamento ou ajuste do modelo (*Training Set*): conjunto de registros usados no qual o modelo é desenvolvido;
2. Conjunto de Testes (*Test Set*): conjunto de registros usados para testar o modelo construído;
3. Conjunto de Validação (*Validation Set*, opcional): conjunto de registros usados para validar o modelo construído;

Essa partição do conjunto inicial de dados em grupos é necessária para que o modelo não se torne dependente de um conjunto de dados específico e, ao ser submetido a outros conjuntos (com valores diferentes dos usados na construção e validação do modelo), apresente resultados insatisfatórios. Esse efeito é chamado de efeito *Bias*. À medida que se aumenta a precisão do modelo para um conjunto de dados específico, perde-se a precisão para outros conjuntos (CAMILO; SILVA, 2009).

Assim, foram testadas várias configurações referentes à distribuição dos dados nos conjuntos, treinamento, teste e validação, assumindo para cada tipo de conjunto de dados o melhor. Embora possam existir diferenças entre as porcentagens de dados, caso se trabalhe com séries de diferentes naturezas, porém, devido ao número de observações de alguns conjuntos de dados, não foi possível aplicar a separação em três conjuntos para todos os tipos de dados.

4.1.1 Determinação dos parâmetros C , σ e B

A fase de ajuste dos modelos SVM aborda também o procedimento de determinar os modelos com os valores ótimos para os parâmetros C e σ , que são de domínio do usuário que está aplicando os métodos. Os valores para tais parâmetros podem variar conforme o número de dados utilizados, natureza e até mesmo a abordagem utilizada, classificação ou regressão, portanto se fez necessário introduzir ao ajuste, o procedimento de determinação dos valores ótimos dos parâmetros C e σ .

Assim, criou-se um *looping* para cada parâmetro, onde se ajustou os modelos variando os valores dos parâmetros C e σ . Primeiramente é iniciado o *looping* para o parâmetro C , variando a partir de $C = 5$, aumentando em 15 unidades a cada final de ciclo, com σ fixado em $\sigma = 5$. Ao se obter o valor para C que minimiza o erro $RMSE$ e maximiza o coeficiente de correlação R , inicia-se o *looping* para o parâmetro σ , com o valor C fixo obtido no ciclo anterior, começando em $\sigma = 5$, sendo somados 15 ao final de cada ciclo.

Em geral, criou-se um laço de até 1000 repetições, assegurando uma amplitude de 15.000 entre o início do *looping* e o final, para cada parâmetro. Ao final do processo de ajuste valores para os parâmetros, uma segunda análise é aplicada. Avaliam-se os valores próximos aos supostos ótimos obtidos, variando os

parâmetros em casas decimais, a fim de refinar e obter os valores ótimos para os parâmetros.

Para o valor B , número de reamostragens *Bootstrap* da amostra original, foram testados valores a partir de 500 reamostras, aumentando a cada novo teste em 100 unidades o valor de B até chegar ao valor máximo de $B = 20.000$. Essa fase de ajuste foi feita posteriormente ao ajuste dos parâmetros C e σ , não prejudicando o ajuste do modelo a partir da amostra original.

Na seção 5 Resultados, serão apresentados apenas os resultados dos modelos com os parâmetros C , σ e B já otimizados, de modo a simplificar a exposição dos dados obtidos e apresentar apenas os valores ótimos ajustados, sendo esses os resultados confrontados com os métodos convencionais já consagrados ou contra os valores obtidos nas literaturas de base.

4.2 MEDIDAS DE AVALIAÇÃO E SELEÇÃO DE MODELOS

4.2.1 Seleção de modelos

Ao se fazer a análise de qual modelo se ajusta aos dados de forma satisfatória, ou ainda, verificar se o modelo é capaz de se adequar aos dados de ajuste, na descrição de uma série temporal, diversos modelos acabaram sendo testados. Assim, pode-se criar uma classificação contendo os melhores modelos utilizando o critério do valor do coeficiente de correlação (R) (TEIXEIRA, 2004):

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y})(y_i^{(p)} - \bar{y}^{(p)})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (y_i^{(p)} - \bar{y}^{(p)})^2}}, \quad (70)$$

onde, $y_i^{(p)}$ é o valor previsto para uma certa observação, y_i é o valor real da observação, $\bar{y}^{(p)}$ e \bar{y} são as médias dos valores das previsões e valores reais das observações, respectivamente.

Assim, assume-se que um modelo é melhor que outro caso seu valor do coeficiente de correlação seja maior, e quanto mais perto de 1 (um), melhor será o modelo ajustado.

4.2.2 Avaliação das previsões

Existem diversos métodos de avaliação a se aplicar ao trabalho desenvolvido. Pode-se analisar o desempenho das previsões, em relações a acertos. Foi essa a avaliação considerada neste trabalho. Sendo que, foram considerados dois métodos descritos a seguir:

i. Raiz do Erro Quadrático Médio (*RMSE – Root Mean Squared Error*)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^{(p)})^2} . \quad (71)$$

Pode-se também utilizar a proporção de predição considerada correta, onde se classifica como corretas as previsões para valores futuros que se encaixem dentro das margens dos vetores suporte. Tais margens são obtidas pelo método *Bootstrap*, assim pode-se definir mais um método de avaliação.

ii. Critério com base nas margens flexíveis

Semelhante ao Erro Médio Absoluto Percentual (MAPE), porém, avalia-se y_i para todo i por:

$$\begin{cases} y_i^{(p)} = 1 \\ y_i^{(p)} = 0 \end{cases} \quad \text{se} \quad \begin{cases} y_i^{(p)} \in I_{bootstrap} \\ \text{caso contrário} \end{cases} \quad (72)$$

e a equação do erro é dada pela mesma expressão:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i^{(p)} - y_i}{y_i} \right) \times 100\% . \quad (73)$$

4.3 ERRO PADRÃO *BOOTSTRAP*

Para calcular as novas margens flexíveis para as previsões foi utilizado o intervalo de confiança *Bootstrap* e para isso, foi necessário calcular primeiramente o erro padrão *Bootstrap*, que foi utilizado no intervalo a ser apresentado na seção seguinte.

O erro padrão *Bootstrap* é calculado para as previsões calculadas pelo método do SVR *Bootstrap*, utilizando todos os B conjuntos de amostras *Bootstrap*. Sua expressão apenas depende dos valores das previsões *Bootstrap* e da média de cada previsão:

$$\hat{e}p_i^* = \sqrt{\frac{1}{B-1} \sum \left(y_j^{i*} - \frac{1}{B} \sum y_j^{i*} \right)^2}, \quad (74)$$

onde B é o número de amostras *Bootstrap* e y_j^{i*} é o valor da previsão *Bootstrap*, com $i = 1, \dots, k$ e $j = 1, \dots, B$.

4.4 INTERVALO DE CONFIANÇA *BOOTSTRAP* – t

O método *Bootstrap*, juntamente com o seu erro padrão, foi utilizado na construção do intervalo de confiança para as previsões feitas pelo método *Support Vector Regression* (SVR).

O intervalo de confiança *Bootstrap-t* é dado pela probabilidade a seguir, utilizando o erro padrão *Bootstrap*:

$$P\left(y_i - z_{1-\alpha/2} \hat{e}p^*(y_i) \leq y_i \leq y_i + z_{1-\alpha/2} \hat{e}p^*(y_i) \right) = 1 - \alpha \quad (75)$$

com n o tamanho da amostra original, $\hat{e}p^*$ é o erro padrão *Bootstrap* da estatística adotada, y_i é o valor previsto utilizando o método SVR convencional para a amostra original, e α é o nível de confiança fixado *à priori*.

Por fim, o intervalo *Bootstrap* para os valores previstos calculados pelo método SVR, utilizando a modelagem LS-SVR é dado por:

$$I_{bootstrap} = \left(y_i - z_{1-\alpha/2} \hat{ep}^*(y_i) \leq y_i \leq y_i + z_{1-\alpha/2} \hat{ep}^*(y_i) \right). \quad (76)$$

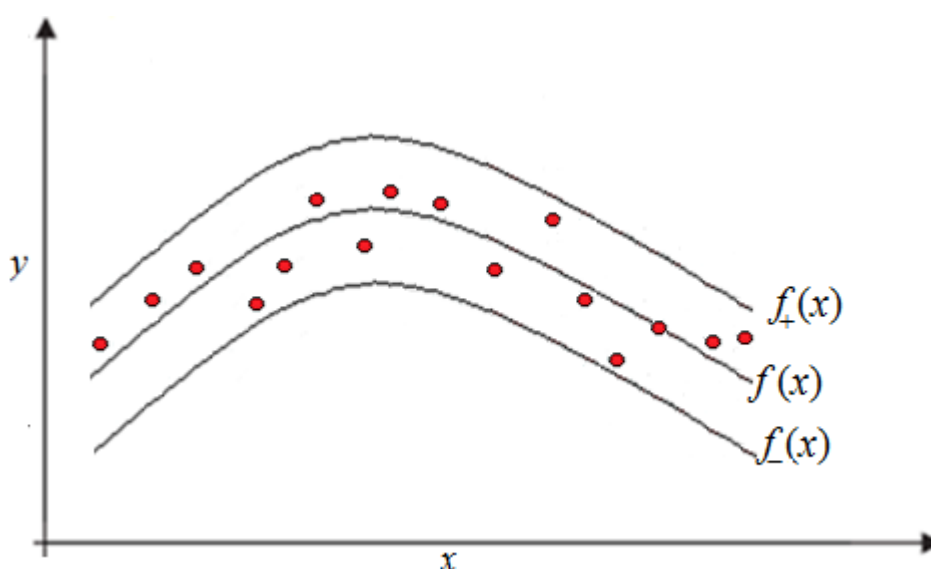
Embora seja flexível e praticamente automático, o cálculo do intervalo de confiança possui um problema que pode afetar a sua eficácia. O método funciona bem quando a distribuição *Bootstrap* da estatística é aproximadamente Gaussiana e a estatística pouco viciada. Sendo respeitadas essas condições, o intervalo de confiança *Bootstrap-t* pode ser calculado na estimação de vários parâmetros.

4.5 MARGENS BOOTSTRAP

O desenvolvimento das margens *bootstrap* é feito com base nos limites do intervalo *Bootstrap-t* calculado para os valores previstos pelo método SVR, determinados pelo LS-SVR, com base no conjunto de dados originais. Assim, após calculados os valores da função preditora, calcula-se os valores das funções preditoras dentro dos conjuntos *Bootstrap* criados, os erros *Bootstrap* e formam-se as novas margens flexíveis *Bootstrap*.

A (FIGURA 12) a seguir apresenta uma exemplificação hipotética para a aplicação das margens *Bootstrap*.

FIGURA 12 – EXEMPLO DE APLICAÇÃO DAS MARGENS *BOOTSTRAP* PARA O MÉTODO SVR



FONTE: O autor (2016).

4.6 METODOLOGIAS PROPOSTAS

Neste trabalho, foram desenvolvidas três metodologias distintas, todas com ênfase em problemas de separações e classificações de padrões e na regressão não linear múltipla, para problemas de previsão.

A primeira baseada no coeficiente de correlação, que exerce o papel de constante reguladora ou penalizadora, e é inserido no início da fase de treinamento, originando o método denominado SVM Correlacionado.

A segunda é uma nova forma de treinamento semi-supervisionado para o método SVM e SVR, com base na replicação *Bootstrap*. Onde se modifica a fase de ajuste dos modelos, utilizando as amostras *Bootstrap* nesta fase, liberando as amostras reais apenas a serem aplicadas na fase de teste.

A terceira metodologia se baseia no intervalo *Bootstrap-t* para gerar uma forma alternativa às margens flexíveis para problemas de regressão. Assim, apresentando uma nova técnica a ser aplicada para a obtenção das margens flexíveis, denominadas de margens flexíveis *Bootstrap*.

As metodologias foram aplicadas a problemas de classificações de padrões e problemas de regressão não linear múltipla. Pode-se destacar o treinamento *Bootstrap*, sendo aplicado principalmente para conjuntos de dados reduzidos, gerando novas amostras para o treinamento e separando as amostras conhecidas para a fase de testes.

Também se introduziu, nas fases de treinamento e teste/validação, a constante criada com base no coeficiente de correlação das variáveis independentes e a variável dependente, evitando o descarte de variáveis independentes menos correlacionadas com a variável dependente. Tais metodologias estão descritas nas seções a seguir.

4.6.1 Treinamento *Bootstrap* para os modelos SVM

Como o método do SVM é um método de aprendizado supervisionado, tem-se a divisão em treinamento e teste/validação. Assim criou-se uma modificação na etapa inicial, etapa de treinamento (ajuste do modelo).

A etapa do treinamento foi modificada, introduzindo o método *Bootstrap*, nos dados de treinamento, a fim de aumentar o número de observações de treinamento,

sendo indicada a sua aplicação principalmente a problemas onde exista um número reduzido de dados. Assim, utilizam-se os conjuntos de amostras *Bootstrap* na fase de treinamento e o conjunto inicial de dados foi separado e usado apenas na fase de teste/validação.

Os ajustes dos modelos SVM foram obtidos pelo método LS-SVM, visando eliminar a complexidade computacional e diminuindo o tempo utilizado com o cálculo dos multiplicadores de Lagrange para cada um dos B conjuntos de treinamento *Bootstrap*, gerando $B \times (n + 1)$ multiplicadores de Lagrange.

Descrição do procedimento do treinamento *Bootstrap*:

- i. A partir do conjunto de n dados originais já conhecidos, divide-se entre as categorias (por exemplo, categoria -1 e categoria +1) que se deseja classificar. Para cada categoria destacou-se os menores e os maiores valores de cada variável preditora, formando intervalos com certa amplitude para cada uma das variáveis independentes;
- ii. Inicia-se a reamostragem *Bootstrap*, mas em vez de fazer os sorteios aleatórios apenas com os valores discretos do conjunto de dados já conhecido, o sorteio foi feito dentro dos intervalos criados na etapa (i), para cada categoria de classificação, e é gerado o mesmo número n de reamostras do conjunto original. Assim, geraram-se B conjuntos diferentes de treinamento, e não apenas B amostras *Bootstrap*. Ao todo foram gerados $B \times n$ amostras *Bootstrap*;
- iii. Atribuíram-se as classes: todas as reamostras geradas a partir dos intervalos formados com os dados da categoria -1 receberam $y_i = -1$, e por sua vez, as reamostras realizadas com os limites coletados dos dados de categoria +1 receberam os valores $y_i = +1$;
- iv. Realiza-se o treinamento do SVM em cada um dos B conjuntos *Bootstrap*, calculou-se os valores ótimos para os multiplicadores de Lagrange pelo método LS-SVM, e geraram-se os B conjuntos de multiplicadores de Lagrange (α) do modelo SVM;

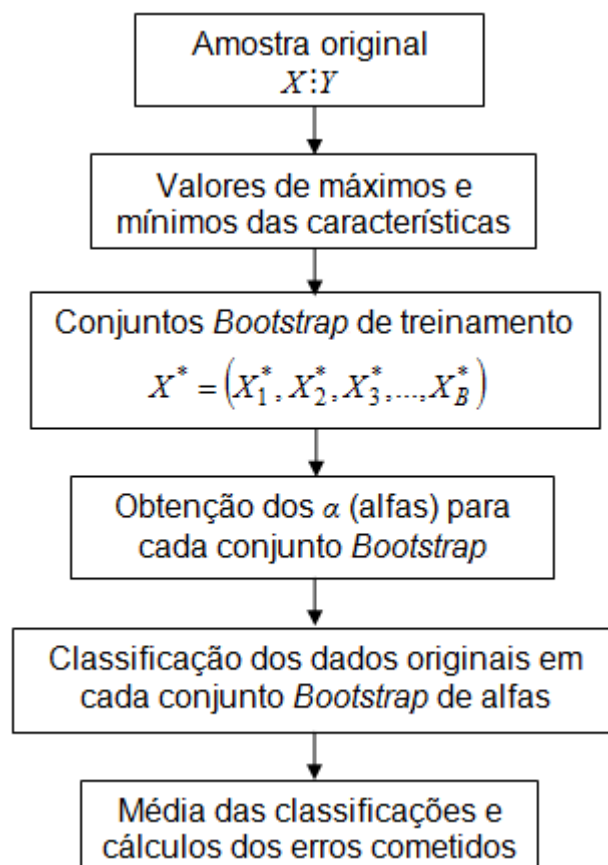
v. Realiza-se a classificação do conjunto original para cada um dos B conjuntos *Bootstrap* dos multiplicadores de Lagrange (α^*), contabilizando cada uma das classificações;

vi. Calcularam-se as médias de cada uma das classificações dos dados originais em relação aos multiplicadores de Lagrange *Bootstrap*, calcularam-se os erros cometidos, pelos métodos já descritos anteriormente;

vii. Ao final, reservou-se a apresentação dos resultados, juntamente com o valor do erro *RMSE* e a porcentagem de erros cometidos, ou seja, padrões +1 classificados como -1 e vice-versa.

A (FIGURA 13) apresenta o diagrama da aplicação do método SVM *Bootstrap* para problemas de classificação de padrões em duas ou mais classes distintas.

FIGURA 13 – ALGORITMO DO TREINAMENTO *BOOTSTRAP* PARA O MÉTODO SVM



FONTE: O autor (2016).

4.6.2 Treinamento *Bootstrap* para os modelos SVR

O treinamento para o modelo de regressão é similar ao treinamento *Bootstrap* para os problemas de classificações de padrões, porém deve-se utilizar o método *Bootstrap* para calcular, também, os valores da variável Y , destacando os valores de máximo e mínimo dentro dos valores da amostra original.

Assim, devem-se destacar todos os intervalos de valores máximos e mínimos para as variáveis, antes de iniciar o processo de reamostragem *Bootstrap*, separando a amostra original para a fase de teste do modelo de regressão. Essa é a principal vantagem em relação ao método SVR convencional, uma vez que, assim se fez possível utilizar todos os dados originais para a fase de teste e auxiliando a verificação da generalização do modelo.

Todas as soluções dos modelos SVR foram obtidas pelo método LS-SVR, visando eliminar a complexidade computacional e diminuindo o tempo utilizado com a solução dos B modelos SVR e o cálculo dos multiplicadores de Lagrange, durante o treinamento *Bootstrap*, onde são gerados $B \times (n + 1)$ multiplicadores de Lagrange $B \times n$ amostras *Bootstrap*.

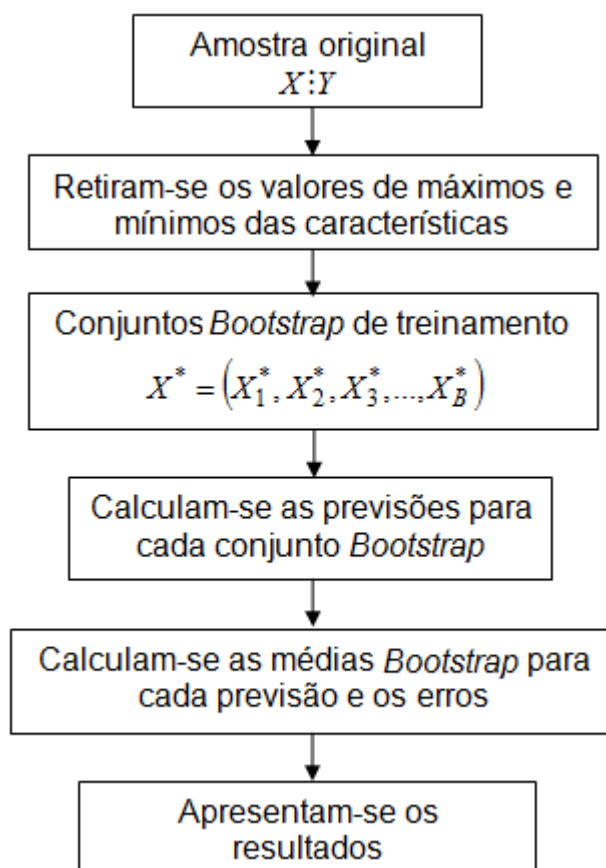
Descrição do procedimento do treinamento *Bootstrap*:

- i. A partir do conjunto de n dados originais já conhecidos, destacam-se os valores máximos e mínimos de cada variável, inclusive a variável a ser prevista, e assim formam-se os intervalos para cada variável com certa amplitude;
- ii. Inicia-se a reamostragem *Bootstrap*, mas em vez de fazer os sorteios aleatórios apenas com os valores discretos do conjunto de dados já conhecido, o sorteio é feito dentro dos intervalos criados na etapa (i), tanto para as variáveis independentes, quanto para a variável a ser prevista, e são gerados o mesmo número n de reamostras do conjunto original, assim, geram-se B conjuntos diferentes de treinamento;
- iii. Realiza-se o treinamento do modelo SVR para cada um dos B conjuntos *Bootstrap*, calculando os valores ótimos para os multiplicadores de Lagrange pelo método LS-SVR;

- iv. As previsões para cada um dos B conjuntos;
- v. Calculam-se as médias de cada uma das previsões dos dados *Bootstrap*, calculam-se os erros cometidos, pelos métodos já descritos anteriormente;
- vii. Apresentam-se os resultados das médias das previsões *Bootstrap*, juntamente com o valor do erro *RMSE*. Pode-se utilizar o processo do calculo das margens *Bootstrap* que estão apresentadas na seção seguinte.

A (FIGURA 14) apresenta o diagrama da aplicação do método SVR *Bootstrap* para problemas de regressão. Neste caso, regressão não linear múltipla.

FIGURA 14 – ALGORITMO DO TREINAMENTO *BOOTSTRAP* PARA O MÉTODO SVR



FONTE: O autor (2016).

4.6.3 Margens flexíveis *Bootstrap* para modelos SVR

Com a inclusão do método *Bootstrap* na etapa de treinamento, criou-se a possibilidade de, após obtenção da função preditora, criar intervalos onde existia uma probabilidade de que os valores reais sejam encontrados, ou seja, a criação de intervalos *Bootstrap* para cada uma das previsões. Assim, esta tese propõe uma modificação na fase de treinamento do SVR, introduzindo a etapa *Bootstrap*, gerando dois modelos distintos, sendo o primeiro o modelo convencional seguindo os passos do método SVR e, o segundo, com método *Bootstrap*, com o segundo mais sensível às variabilidades existentes nas previsões.

Essa probabilidade pode ser dada *à priori*, sendo assim, possível de considerar probabilidades mais, ou menos, rígidas, criando mais um parâmetro de comparação ao método SVR convencional, e substituindo as margens originais do método SVR, onde se caracterizam por serem basicamente “paralelas”, ou virtualmente flexíveis, por novas margens que são realmente flexíveis. Para a solução desse método utilizou-se o método LS-SVR para facilitar o cálculo dos multiplicadores de Lagrange, uma vez que, ao todo, são gerados $B \times (n + 1)$ multiplicadores de Lagrange.

Esse método pode ser associado ao treinamento *Bootstrap* apresentado na seção anterior, levando em consideração a fase de treinamento do método SVR *Bootstrap*.

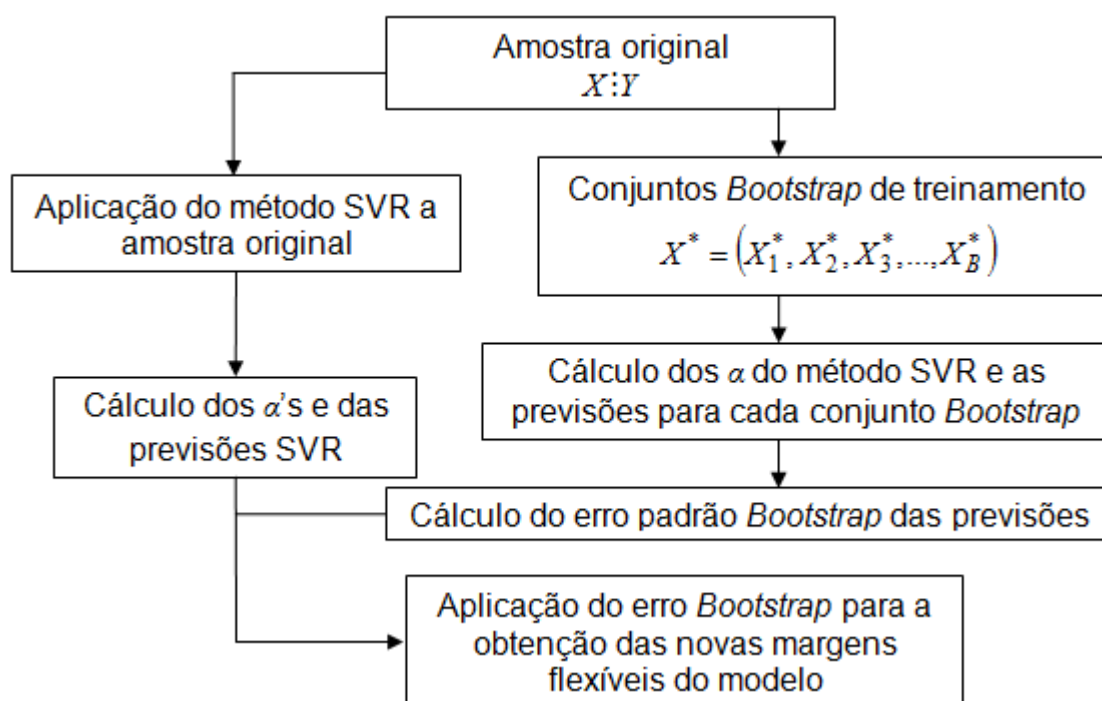
Descrição do procedimento do treinamento *Bootstrap*:

- i. Calculam-se os multiplicadores de Lagrange e as previsões pelo método SVR para a amostra original;
- ii. A partir da amostra inicial, inicia-se a reamostragem *Bootstrap*, gerando o mesmo número n de reamostras do conjunto original, assim, geram-se B conjuntos diferentes de treinamento, e não apenas B amostras *Bootstrap*;
- iii. Atribuem-se as os valores da variável Y da amostra original a cada uma das reamostras *Bootstrap* seguindo a mesma sequência da amostra original;

- iv. Realiza-se o treinamento do SVR em cada um dos B conjuntos *Bootstrap*, gerando os B conjuntos dos multiplicadores de Lagrange do modelo SVR, e os B conjuntos de previsões *Bootstrap*;
- v. Calcula-se a média de cada uma das previsões *Bootstrap*, calcula-se o erro padrão *bootstrap* e criam-se os limites do intervalo *Bootstrap*, identificando os limites inferiores e superiores para cada uma das previsões, gerando assim os limites das margens *bootstrap* para o modelo SVR;
- vi. Destacam-se os valores das bordas das margens *Bootstrap*, unindo os valores previstos pelo método SVR comum, calculam-se os erros do SVR e verifica-se se todas as previsões se encontram dentro das margens flexíveis *Bootstrap*, demonstrando a robustez do modelo desenvolvido.

A (FIGURA 15) a seguir apresenta o fluxograma da aplicação *Bootstrap* ao método SVR, gerando a modificação na etapa de treinamento e expressando a metodologia das margens flexíveis *Bootstrap* para modelos SVR:

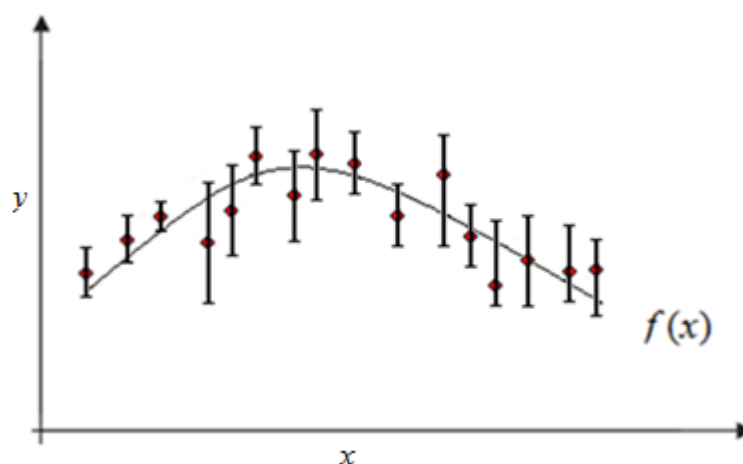
FIGURA 15 – ALGORITMO DAS MARGENS FLEXÍVEIS *BOOTSTRAP* PARA O MODELO SVR



FONTE: O autor (2016).

A (FIGURA 16) a seguir apresenta uma exemplificação da determinação dos valores das previsões, bem como os limites *Bootstrap* para cada uma das previsões. Para cada valor previsto, é criado um intervalo, simétrico, centralizado na previsão e com as novas bordas obtidas pelo método *Bootstrap* descrito na (FIGURA 15). Tais limites são calculados com base nos multiplicadores de Lagrange, obtidos pelo método SVR.

FIGURA 16 – INTERVALO DE VARIAÇÃO PARA AS PREVISÕES FEITAS PELO SVR *BOOTSTRAP*



FONTE: O autor (2016).

4.6.4 SVM e SVR Correlacionado

Essa metodologia de treinamento e teste consiste em calcular a correlação entre as variáveis independentes e as classes dos dados, ou os valores da variável de previsão. Basicamente, antes de efetuar o treinamento, calcula-se a correlação das variáveis independentes e, assim, modifica-se o conjunto de treinamento, destacando as variáveis que possuem uma maior correlação com a variável independente, porém, mantendo-se as variáveis de menor correlação no conjunto de treinamento e teste/validação.

Verificando a aplicação da metodologia aos problemas de classificação e aos problemas de regressão. É importante ressaltar que tal modificação não altera o valor absoluto da correlação entre as variáveis. A mudança ocorre caso a correlação seja negativa, que passa a ser positiva. O conjunto de teste/validação também deve ser multiplicado pela correlação.

Tal modificação pode ser utilizada combinada com as metodologias apresentadas anteriormente, desde que os resultados obtidos, no mínimo, sejam semelhantes às metodologias descritas. Ela deve ser inserida antes da aplicação do *Bootstrap*, resultando em uma fase pré-ajuste, não interferindo diretamente na fase de teste.

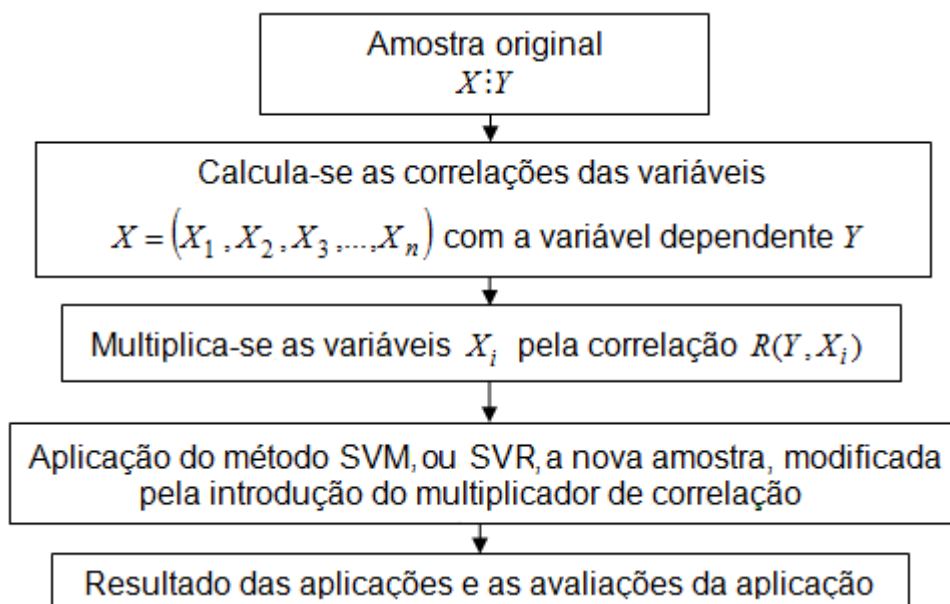
Todas as aplicações dessa modificação foram confrontados com as aplicações convencionais e seus resultados foram comparados a fim de atestar a melhora dos resultados.

Descrição dos métodos SVM e SVR Correlacionados:

- i. Calculam-se as correlações entre todas as variáveis independentes e a variável dependente, tanto para problemas de correlação, como de classificação, utilizando todos dados existentes;
- ii. Multiplicam-se as variáveis independentes pela sua respectiva correlação em relação a variável dependente;
- iii. Identifica-se o método a ser utilizado, caso o método SVM/SVR *Bootstrap* seja o escolhido, efetua-se a multiplicação da correlação aos limites destacados para o sorteio das amostras *Bootstrap*;
- iv. Aplicam-se os modelos, SVM ou SVR, aos novos dados já multiplicados pela correlação, otimizando os parâmetros C e σ ;
- v. Apresentam-se os dados da aplicação.

A (FIGURA 17) a seguir apresenta o fluxograma da aplicação do método SVM/SVR Correlacionado, levando em consideração a nova constante penalizadora com base no coeficiente de correlação, gerando uma fase pré-ajuste do modelo, não modificando a etapa de ajuste:

FIGURA 17 – TREINAMENTO COM A INTRODUÇÃO DO MULTIPLICADOR DE CORRELAÇÃO



FONTE: O autor (2016).

4.7 DADOS DAS APLICAÇÕES REAIS

4.7.1 Amostras *Bootstrap*

Para a aplicação do método *Bootstrap* para o treinamento é necessário considerar alguns critérios referentes ao conjunto onde se deve fazer o sorteio aleatório. Isso se deve à natureza dos dados utilizados, onde o período, hora, dia ou mês, pode influenciar no resultado, por exemplo, a estação meteorológica do período pode influenciar no tempo total de exposição do sol, que por sua vez modifica o nível de radiação solar que atinge a área estudada, ou até mesmo podendo modificar o ecossistema estudado, entre outros.

Assim, para cada tipo de aplicação, foi levada em consideração a criação de subgrupos para as amostras, mudando assim o algoritmo de criação das amostras *Bootstrap*, aplicando a replicação *Bootstrap* para cada um desses subgrupos formados a partir das amostras iniciais.

Tal subdivisão está apresentada nas seções a seguir, sendo considerado cada tipo de problema, com o seu específico processo de subdivisão específico, podendo ser generalizado apenas a amostras de mesma natureza de dados.

4.7.2 Dados dos exames da Ceratocone

A córnea possui basicamente cinco camadas, sendo a última delas o endotélio, responsável pela nutrição e metabolismo da córnea. As células dessa camada são em forma de hexágono e com tamanhos parecidos. Para que ele exerça sua função normal, é necessário que as células tenham densidade, tamanho e formas dentro do normal.

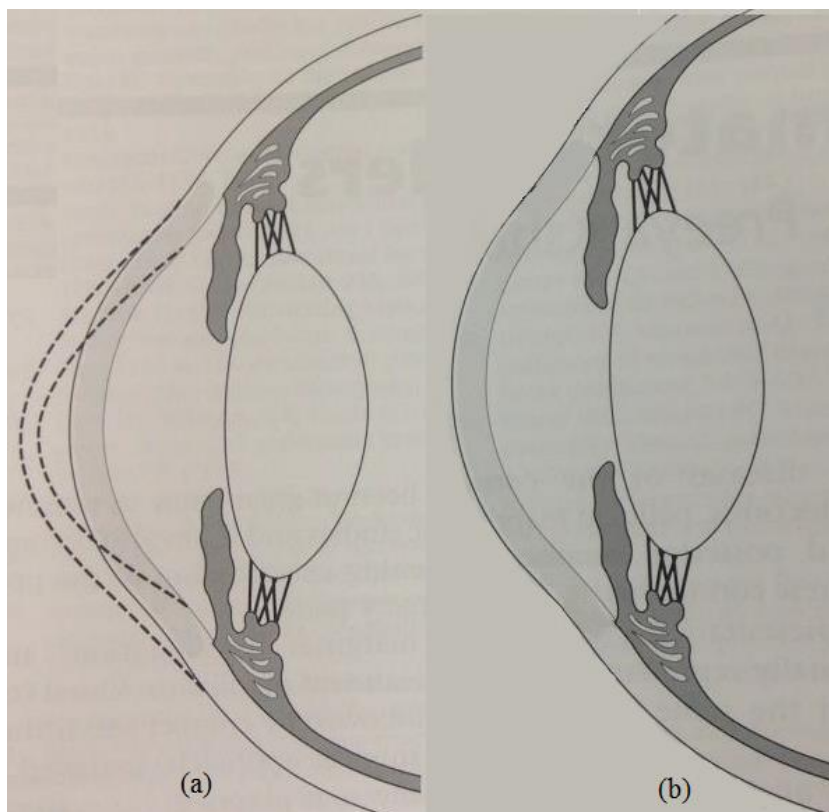
O Ceratocone é uma doença degenerativa que atinge a córnea, não inflamatória, bilateral e assimétrica, progressiva, levando às inúmeras alterações na superfície da córnea. Caracteriza-se por afinamento central, protusão apical e astigmatismo irregular, com vários graus de cicatrização, ocasionando uma redução da acuidade visual. A córnea adquire forma cônica devido ao seu afinamento e protusão. Não existe infiltração celular ou vascularização. O cone pode ser redondo ou oval, podendo localizar-se próximo ao eixo visual, superior ou inferior a ele (ELIAS *et al*, 2005).

O Ceratocone também é caracterizado pela presença de astigmatismo irregular, aumento da curvatura da córnea, afinamento e cicatriz apical. O afinamento e a protrusão da córnea originam deslocamento apical responsável pelo astigmatismo irregular (FELICIANO DE DEUS, 2015; MOREIRA; MOREIRA, 2000).

A incidência média do Ceratocone é de dois novos casos para cada 100.000 indivíduos ao ano e sua prevalência tem reportado com frequência de 1:435 à 1:2000. Dados estatísticos mostram que a incidência do ceratocone é de 1:4000 asiáticos e 1:30.000 europeus que vivem no Reino Unido (FELICIANO DE DEUS, 2015; BOYD, 2012).

A (FIGURA 18) apresenta o principal efeito que a doença Ceratocone causa na córnea (FIGURA 18a) e uma córnea sadia, sem o efeito do Ceratocone (FIGURA 18b).

FIGURA 18 – OLHO AFETADO PELA CERATOCONE (a) E OLHO NORMAL (b)



FONTE: Adaptado de FELICIANO DE DEUS (2015).

Os exames feitos apresentam algumas características como idade e sexo, que não entraram no modelo e os valores das variáveis independentes, que essas sim, fizeram parte da modelagem do SVM: K central, KISA%, Coeficiente de variação, densidade e Hexagonalidade, sendo descritos a seguir (FELICIANO DE DEUS, 2015; SOUZA, 2011):

i. O KISA% é um valor derivado do produto de 4 índices: A curvatura corneana central (K central), o I-S (quantidade de inclinação da parte inferior da córnea em comparação com o da córnea superior), um índice refletindo o astigmatismo regular (IRAR) e outro índice para o astigmatismo irregular (IAI), e é dado pela equação;

$$KISA\% = (K \text{ central}) \cdot (I - S) \cdot (IRAR) \cdot (IAI) \cdot \frac{100}{300}; \quad (77)$$

ii. Coeficiente de Variação (CV): Polimegatismo (aumento no tamanho das células) valores maiores que 30% são considerados ruins, pois caracteriza o aumento celular detectado nas células endoteliais;

iii. Densidade: Densidade Celular medida por células/mm², valores abaixo do intervalo 2400-3500 cell/mm² são considerados ruins para o funcionamento celular;

iv. Hexagonalidade: Pleomorfismo (mudança na forma das células), valores acima de 60% são esperados. Valores abaixo de 60% significam perda significativa da forma celular.

Esse conjunto de dados engloba quatro classes referentes aos níveis, ou grau, da doença em que a córnea é classificada. Ao todo, vinte e quatro observações de pacientes classificados com grau 1, vinte e um pacientes com grau 2, quatro pacientes de grau 3 e cinco pacientes de grau 4, apresentando uma disparidade entre os graus 3 e 4 e os graus 1 e 2. Assim, o método SVM *Bootstrap* foi utilizado diretamente nas classificações para o 3 contra o grau 4, por se possuir um número reduzido de amostras de exames.

A (TABELA 1) apresenta um resumo da ficha a ser preenchida com os valores dos exames médicos feitos.

TABELA 1 – FICHA DOS EXAMES (CERATOCONE)

ID PX	OLHO	IDADE	SEXO	K CENTRAL	KISA%	GRAU	% (CV)	DENSIDADE	% HEXA.
4669	E	16	F	68,07	20346%	IV	30.8%	3008.7	65%
3517	E	40	M	58,08	18715%	IV	43.6%	3823.0	49%
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

FONTE: O autor (2016).

A fim de deixar a leitura do trabalho mais fluida, os dados estão descritos de forma detalhada na seção de anexos do trabalho (ANEXO 1), restando ao corpo linear da tese apenas os resultados coletados das aplicações feitas e seus devidos processos de avaliação de desempenho.

4.7.3 Dados do reservatório de Manasgaon

Devido à dificuldade de se obter dados de reservatórios nacionais foram utilizados dados da literatura. Esses dados correspondem à reserva de água de Manasgaon, reservatório situado em Anand Sagar, Shegaon, Índia (DESWAL; PAL, 2008). Para as previsões utilizou-se um conjunto de dados composto de 48 observações do nível de evaporação referentes à média semanal da reserva de Manasgaon, sendo os padrões de entrada separados em quatro características independentes: média da temperatura ($^{\circ}\text{C}$), velocidade média do vento (m/s), horas de sol (h/dia) e média da umidade relativa do ar (%), e o padrão dependente: evaporação, por milímetros por semana.

Para esse conjunto de dados, visando comparar os resultados obtidos com os da literatura do conjunto original de 48 observações foram separadas 34 amostras para a fase de treinamento e as 14 restantes para a fase de teste/validação. Para esse conjunto de dados, foi aplicada a metodologia das novas margens flexíveis *Bootstrap* aliada à introdução da nova constante baseada no coeficiente de correlação.

Ao se aplicar o método *Bootstrap*, também tem-se o cuidado em separar os dados conforme as estações meteorológicas para que os dados, por exemplo, verificou-se que a diferença entre os valores coletados no verão são, estatisticamente, diferentes que os dados coletados no inverno, influenciando assim as amostras *Bootstrap*.

Assim como para os dados dos exames da Ceratocone, a descrição completa dos dados está apresentada nos anexos deste trabalho (ANEXO 2), visando o dinamismo da apresentação dos resultados.

4.7.4 Dados do exame do grau de astigmatismo

Este conjunto de dados apresenta os resultados dos exames para a avaliação do grau do astigmatismo dos olhos. Composto pelo exame feito em 26 pacientes, todos com grau de pelo menos 0,25 de astigmatismo, os exames ainda apresentam algumas observações qualitativas, tais como idade, sexo, profissão e olho afetado, que não entraram na modelagem do problema.

A princípio, as informações referentes à idade dos pacientes constavam na modelagem do problema, porém, como não gerou ganhos consideráveis com a inclusão, optou-se por retirar a variável idade nos modelos finais, cujos resultados estão apresentados neste trabalho.

Para esta aplicação considerou-se o valor do astigmatismo como sendo a variável dependente a ser prevista com base em quatro variáveis independentes, sendo elas: Ceratometria, subdividida em Eixo Mais Plano e Eixo Mais Curvo; e Refração, subdividido em Esfera e Cilíndrico.

A (TABELA 2) apresenta a ficha de cadastro dos pacientes e os valores dos resultados dos pacientes submetidos aos exames.

TABELA 2 – FICHA DOS EXAMES (ASTIGMATISMO)

ID PX	SEXO	IDADE	PROFISSÃO	OLHO AFETADO	Ceratometria		Refração	
					MAIS AMPLO	MAIS CURVO	ESFERA	CILÍND.
1	M	34	OPER. PROD.	OE	41,25	42,5	4,25	-1,75
2	M	71	APOSENTADO	OD	43,75	47	5,25	-2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

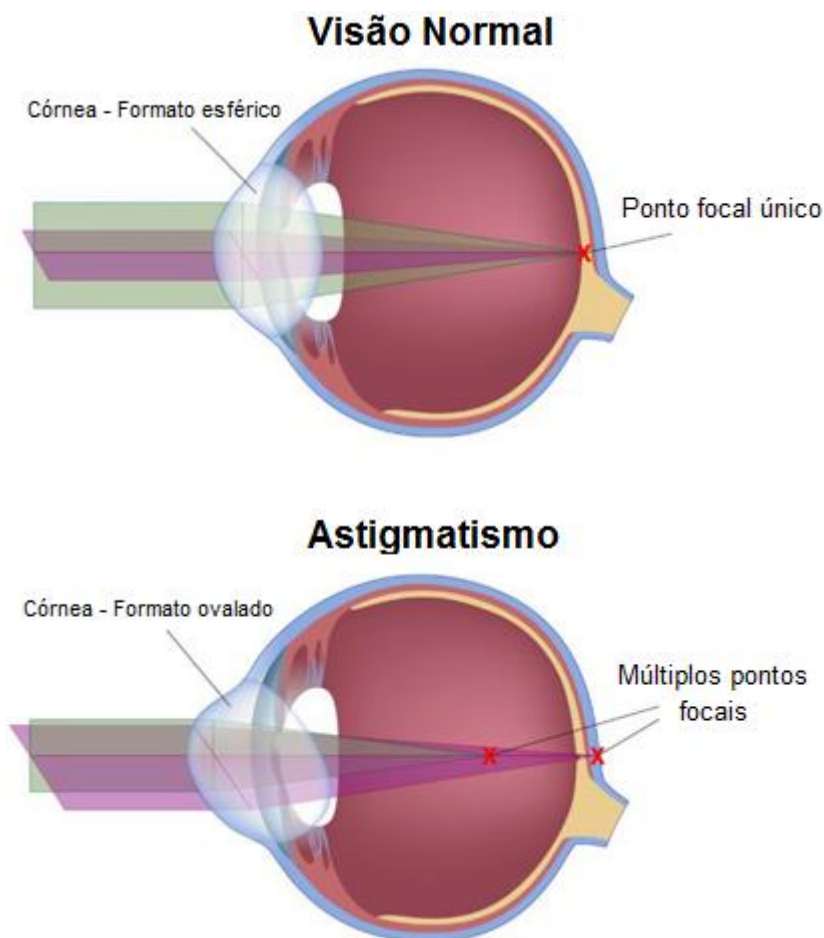
FONTE: O autor (2016).

Para esse conjunto de dados foi aplicado a penalização pela correlação para obter as previsões para os graus de astigmatismo e o treinamento *bootstrap* para gerar as novas margens flexíveis *Bootstrap*.

Os valores discretizados dos exames de todos os pacientes utilizados no desenvolvimento do trabalho antes de serem particionados em grupo de treinamento e grupo de teste estão apresentados de forma completa e detalhados em anexo ao trabalho (ANEXO 3).

A (FIGURA 19) apresenta a diferença entre a incidência da imagem em um olho sadio, ou seja, sem presença de grau astigmatismo, e um olho com certo grau de astigmatismo, exemplificando como a deformidade na córnea humana influência na distorção da imagem projetada no fundo do globo ocular.

FIGURA 19 – DIFERENÇA ENTRE OLHO NORMAL E OLHO COM ASTIGMATISMO



FONTE: O autor (2016).

4.7.5 Dados meteorológicos

Para esse tipo de dados, as observações foram coletadas no portal INMET Instituto Nacional de Meteorologia, de livre acesso à base de dados.

Uma estação meteorológica automática (EMA) coleta, de minuto em minuto, as informações meteorológicas (temperatura, umidade, pressão atmosférica, precipitação, direção e velocidade dos ventos, radiação solar) representativas da área em que está localizada. A cada hora, estes dados são integralizados e disponibilizados para serem transmitidos, via satélite ou telefonia celular, para a sede do INMET, em Brasília. O conjunto dos dados recebidos é validado, através de um controle de qualidade e armazenado em um banco de dados (REDE, 2016).

A coleta de dados é feita através de sensores para medição dos parâmetros meteorológicos a serem observados. As medidas tomadas, em intervalos de minuto a minuto, e integralizadas para no período de uma hora, para serem transmitidas, são (REDE, 2016):

- i. Temperatura Instantânea ($^{\circ}\text{C}$);
- ii. Umidade Relativa Instantânea do Ar (%);
- iii. Temperatura do Ponto de Orvalho;
- iv. Pressão Atmosférica Instantânea do Ar (hPa);
- v. Velocidade do Vento (m/s);
- vi. Radiação Solar (kJ/m^2);
- vii. Precipitação acumulada no período (mm).

Parâmetros meteorológicos são descritos como:

- i. Temperatura instantânea em $^{\circ}\text{C}$: valor da temperatura do ar em $^{\circ}\text{C}$ no momento da medição;
- ii. Umidade relativa instantânea do ar (%): valor da umidade relativa do ar em % no momento da medição;
- iii. Temperatura do ponto de orvalho: Se refere à temperatura a partir da qual o vapor d'água contido na porção de ar de um determinado local sofre condensação. Quando a temperatura do ar está abaixo do ponto de orvalho, normalmente tem-se a formação de névoa seca ou neblina;
- iv. Pressão Atmosférica (hPa): É a força por unidade de área, exercida pelo ar contra uma superfície. A pressão atmosférica é medida por meio de um equipamento conhecido como barômetro. Essas diferenças de pressão têm uma origem térmica estando diretamente relacionadas com a radiação solar e os processos de aquecimento das massas de ar. Formam-se a partir de influências naturais, como: continentalidade, maritimidade, latitude, altitude, etc.;

v. Velocidade do vento (m/seg): Registro da velocidade (m/s) e rajada (m/s) de hora em hora. Resulta do deslocamento de massas de ar, derivado dos efeitos das diferenças de pressão atmosférica entre duas regiões distintas e é influenciado por efeitos locais como a orografia e a rugosidade do solo;

vi. Radiação (em kJ/m^2): Radiação solar é a designação dada à energia radiante emitida pelo Sol. A radiação solar fornece anualmente para a atmosfera terrestre cerca de $1,5 \times 10^{18}$ kWh de energia;

vii. Chuva (mm): Registro automático em mm do total de chuva de hora em hora.

Para esse conjunto de dados, primeiramente foi fixada a variável dependente, temperatura, e as demais como sendo as variáveis independentes, somando assim seis variáveis independentes. Foi considerado o método de treinamento *Bootstrap* para o *Support Vector Regression*, gerando as previsões de temperatura e as margens flexíveis *Bootstrap*, e os resultados estão apresentados nas seções a seguir, e o conjunto de dados utilizado está listado na seção dos anexos do trabalho (ANEXO 4).

4.7.6 Cotação do Dólar

A oscilação do valor do dólar sofre influência, principalmente, de variáveis qualitativas, por exemplo, desempenho da economia local, previsão do PIB, crises financeiras, especulações e até a própria influência do governo.

Assim, aqui se considerou apenas as variáveis quantitativas, a fim de desenvolver um modelo com base apenas na parte determinística dos dados. Para isso, utilizou-se diretamente dos valores absolutos da cotação ao final de cada dia, ou seja, o valor de fechamento do dia, a partir do dia 03 de novembro de 2015, até o dia 25 de fevereiro de 2016.

Porém, dentro desse intervalo de tempo houveram diversos feriados ou dias em que não houve cotação. Para esses dias, utilizou-se o valor do dia anterior, uma vez que, se não houve cotação no dia, ela se manteve estável até o dia seguinte.

Além dos valores absolutos das cotações, a variação média da semana também foi considerada como variável independente que entrou no ajuste do

modelo. Assim as variáveis independentes se resumem aos valores do fechamento de segunda a quinta, reservando a cotação de sexta para a variável dependente a ser prevista. As variáveis possuem a seguinte configuração:

$$x = (x_{segunda}, x_{terça}, x_{quarta}, x_{quinta}, x_{média}), \quad (78)$$

onde os quatro primeiros valores são referentes às cotações de segunda a quinta e o $x_{média}$ refere-se à média das variações dentro desse dias, e por fim, para os valores de y utilizou-se a cotação do fechamento das sextas-feiras.

Os dados utilizados foram coletados no sitio eletrônico “Uol Economia” sendo a última coleta feita no dia 28 de fevereiro de 2016, porém, as análises e aplicações das metodologias desenvolvidas neste trabalho só levaram em consideração os valores obtidos até o dia 25 de fevereiro de 2016. Como o acesso aos dados é livre, não se considerou necessário a inclusão dos mesmos ao corpo do trabalho.

Para essa aplicação, se utilizou o método SVR *Bootstrap*, que se destacam os valores máximos e mínimos de cada variável independente para fazer o sorteio aleatório *Bootstrap*.

5 RESULTADOS

Nesta seção estão apresentados os resultados obtidos pelas aplicações das metodologias propostas. Para cada aplicação, foi descrita de forma detalhada seu processo desde o treinamento até os resultados finais, e, se existir, a comparação com os resultados coletados na literatura consultada.

Em todas as aplicações, foram descritos os valores para os parâmetros do SVM e *Bootstrap*, uma vez que na fase do treinamento devem-se determinar os valores das constantes C , σ e B que minimizam os erros cometidos, e tais valores foram utilizados na fase de teste/validação. A fase de validação pode ser omitida quando o número de observações for reduzido, impedindo a separação e podendo comprometer a fase de treinamento.

Em todas as aplicações, tanto para problemas de classificação, como para os problemas de previsão, os valores dos multiplicadores de Lagrange foram obtidos pela solução dos modelos LS-SVM (*Least Squares Support Vector Machine*), devido a sua baixa complexidade computacional e a sua implementação mais fácil. Pode-se destacar também que o método LS-SVM apresenta resultados equivalentes ao método SVM tradicional (SANTOS, 2013).

5.1 CERATOCONE

Esta aplicação levou em consideração a classificação referente ao grau da doença que atinge a córnea humana. Assim, em todas as abordagens utilizou-se o método SVM para classificação. Primeiramente utilizou-se o método SVM Correlacionado para separar os conjuntos entre os graus 1 e 2 (conjunto +1) e graus 3 e 4 (conjunto -1), como sendo a aplicação inicial do método.

Ao se classificar dentro dos grandes conjuntos, utilizou-se novamente do método SVM Correlacionado para classificar, no conjunto +1, os padrões referentes aos graus 1 ou 2, uma vez que esse conjunto possuía ao todo 45 amostras dos exames da córnea, não sendo necessário o processo de reamostragem *Bootstrap*.

Devido ao número reduzido de amostras, a aplicação ao grande conjunto -1, conjunto composto por pacientes com graus 3 e 4, se fez necessário a introdução do treinamento *Bootstrap* para se obter uma maior generalização das classificações. Assim utilizou-se o método SVM *Bootstrap* para o treinamento semi-supervisionado,

gerando os B conjuntos *Bootstrap* de amostras sintéticas baseadas nos limites pré-fixados no início do treinamento, e, com isso, liberando as nove amostras conhecidas para a fase de teste.

As considerações a respeito dos parâmetros utilizados no treinamento e teste dos modelos foram dadas nas seções referentes a cada aplicação, bem como o desenvolvimento dos métodos. As aplicações estão apresentadas nas seções a seguir.

5.1.1 Separação entre os graus (1, 2) e (3, 4) pelo método SVM Correlacionado

O conjunto sobre os dados da Ceratocone está dividido em quatro graus diferentes, sendo que os dois primeiros (grau 1 e 2) possuem um número maior (vinte e quatro e vinte e um, respectivamente) que os graus 3 e 4 (quatro e cinco, respectivamente). Para essa classificação, o principal problema é o baixo número de amostras dos graus 3 e 4, em relação aos graus 1 e 2. Para análise, utilizou-se o método SVM Correlacionado, comparando com a porcentagem de acerto obtido pelo método do discriminante linear de Fisher.

A primeira aplicação do SVM foi feita para separar os dados em dois grandes conjuntos: conjunto +1 para os graus 1 e 2; conjunto -1 para os graus 3 e 4, onde apenas a multiplicação pela correlação foi aplicada. Para tal aplicação os resultados obtidos pelo método SVM Correlacionado foram comparados com os resultados do método SVM convencional e Discriminante linear de Fisher, a fim de se verificar a robustez do método desenvolvido.

Para o método SVM Correlacionado obtiveram-se os valores ótimos $C = 26,0$ e $\sigma = 51,5$ para os parâmetros, gerando um erro $RMSE = 0,28991$. Para o método SVM obtiveram-se os valores ótimos $C = 150$ e $\sigma = 235,0$, gerando o erro $RMSE = 0,30208$, apresentando melhora em relação ao método usual. Em ambas as aplicações, obteve-se um acerto de 100% nas classificações. Visando uma simples comparação, utilizou-se também a classificação pelo método do discriminante Linear de Fisher, obtendo uma classificação correta em 19 das 22 amostras de teste, gerando um acerto de 86,4%.

Utilizou-se 32 amostras para o conjunto de ajuste (15 amostras do grau 1, 11 amostras do grau 2, 3 amostras do grau 3 e 3 amostras do grau 4), e as demais 22 amostras foram utilizadas para a fase de teste.

A (TABELA 3) apresenta os valores obtidos pelas classificações feitas pelo método SVM Correlacionado e SVM convencional para a separação entre os dois grandes grupos, classe +1 formada pelos graus 1 e 2, e classe -1 formada pelos graus 3 e 4.

TABELA 3 – CLASSIFICAÇÃO ENTRE OS GRUPOS (1 E 2 CONTRA 3 E 4)

Paciente	SVM Correlacionado	SVM	Categoria	Distância A	Distância B
1	0,9460	0,9444	+1	0,0540	0,0556
2	1,0374	1,0366	+1	0,0374	0,0366
3	0,7298	0,7334	+1	0,2702	0,2666
4	1,1549	1,1311	+1	0,1549	0,1311
5	0,9195	0,9241	+1	0,0805	0,0759
6	1,0230	1,0150	+1	0,0230	0,0150
7	1,1660	1,2135	+1	0,1660	0,2135
8	1,0474	1,0804	+1	0,0474	0,0804
9	1,0284	1,0039	+1	0,0284	0,0039
10	1,0010	1,0173	+1	0,0010	0,0173
11	0,9833	0,9756	+1	0,0167	0,0244
12	0,9436	0,9857	+1	0,0564	0,0143
13	0,8696	0,8767	+1	0,1304	0,1233
14	1,0437	1,0999	+1	0,0437	0,0999
15	0,1609	0,1293	+1	0,8391	0,8707
16	0,7992	0,8191	+1	0,2008	0,1809
17	0,4804	0,4978	+1	0,5196	0,5022
18	1,0231	1,0257	+1	0,0231	0,0257
19	0,9419	0,9757	+1	0,0581	0,0243
20	-0,7143	-0,6076	-1	0,2857	0,3924
21	-0,9635	-0,9647	-1	0,0365	0,0353
22	-0,2342	-0,2056	-1	0,7658	0,7944

FONTE: O autor (2016).

Os valores associados às Distâncias A e B da (TABELA 3) referem-se a diferença entre os valores obtidos pelo método SVM Correlacionado e SVM convencional e as classes corretas, respectivamente, de cada paciente utilizado na fase de teste dos métodos utilizados.

Os valores das correlações entre as variáveis independentes e a variável dependente são apresentados na (TABELA 4). Os valores das correlações foram obtidas com base em todos os dados do modelo, ou seja, foram calculadas em um momento anterior à partição do conjunto inicial de dados.

Tais valores constituem a constante penalizadora baseada na correlação que é proposta pela metodologia SVM Correlacionado. Devem-se multiplicar os dados de entrada antes do treinamento do modelo SVM. Assim, o produto foi utilizado para o ajuste do modelo, e os seus resultados foram confrontados com os obtidos com o método SVM convencional e a classificação feita pelo discriminante linear de Fisher.

TABELA 4 – CORRELAÇÕES ENTRE AS VARIÁVEIS (CERATOCONE)

	K Central (D)	KISA	CV	Densidade	Hexa.
Categoria	-0,7798	-0,4177	0,0201	-0,2577	-0,0026

FONTE: O autor (2016).

Efetuando a classificação pelo método do discriminante de Fisher, os resultados foram piores. Pelo discriminante, houve três erros na classificação, dois pacientes que, na verdade, pertenciam ao grupo +1 (graus 1 ou 2) ficaram classificados no grupo -1 (graus 3 e 4), e um paciente que pertencia ao grupo -1 foi classificado como sendo do grupo +1, gerando assim 3 erros em 22 padrões testados, cerca de 13,64% de erro cometido, evidenciando a maior robustez da metodologia desenvolvida, em relação a métodos já consolidados.

A (TABELA 5) apresenta um resumo do desempenho dos métodos utilizados na aplicação da separação dos graus 1 e 2 contra 3 e 4, evidenciando o melhor desempenho do método SVM Correlacionado em relação ao método SVM convencional e ao método do Discriminante Linear de Fisher.

A (TABELA 5) apresenta o resumo dos resultados obtidos pelos métodos SVM Correlacionado, SVM e Discriminante Linear de Fisher, apresentando os parâmetros ótimos obtidos nos ajustes dos modelos, o erro *RMSE*, coeficiente de correlação *R* e o erro *MAPE* para os métodos.

TABELA 5 – RESUMO DOS PARÂMETROS E ERROS (1 E 2 CONTRA 3 E 4)

Método	<i>C</i>	σ	<i>RMSE</i>	<i>R</i>	<i>MAPE</i>
SVM Correlacionado	26,0	51,5	0,28991	0,90845	100%
SVM	150,0	235,0	0,30208	0,90013	100%
Discriminante linear de Fisher	-	-	-	-	86,36%

FONTE: O autor (2016).

5.1.2 Separação entre graus 1 e 2 pelo método SVM Correlacionado

Para esta aplicação foi utilizado o método SVM Correlacionado, uma vez que o conjunto de dados iniciais somam ao todo 45 pacientes. Os resultados foram confrontados com a aplicação do SVM convencional e com o discriminante de Fisher. Foi considerado a categoria +1 como sendo os pacientes classificados com grau 1 e categoria -1 para os pacientes classificados como grau 2.

Do conjunto inicial de pacientes, separou-se 27 amostras para o treinamento (quinze amostras de grau 1 e doze de grau 2) e 18 para a fase de teste (nove de cada grau), e esta configuração foi repetida para a aplicação do SVM convencional e do discriminante de Fisher.

A (TABELA 6) a seguir apresenta os valores da correlação entre a variável Y e as variáveis independentes dos pacientes, tanto para o grau 1 quanto para o grau 2, que foram utilizadas para o treinamento e teste do método SVM. Não se pode reutilizar a (TABELA 5) das correlações devido a retirada dos dados dos pacientes que estão classificados com grau 3 e 4 da doença, uma vez que os dados desses pacientes foram utilizados em outra aplicação, utilizando o método SVM *Bootstrap*, a ser apresentado nas seções a seguir.

TABELA 6 – CORRELAÇÕES PARA OS DADOS DE GRAU 1 E 2

	K Central (D)	KISA	CV	Densidade	Hexa.
Categoria	-0,8141	-0,3886	-0,3283	-0,2237	0,2185

FONTE: O autor (2016).

Fixando os valores de $C = 4585,0$ e $\sigma = 322,5$, resultando em um erro $RMSE = 0,5730$, com 100% de acerto nas classificações para o método SVM Correlacionado. Para o método SVM comum o erro ficou em $RMSE = 0,6577$, conseguindo, também, classificar todos os pacientes de forma correta, gerando um erro percentual de 0%.

A (TABELA 7) apresenta os valores obtidos pelo método SVM Correlacionado para a separação entre os graus 1 e 2.

TABELA 7 – CLASSIFICAÇÕES PARA PACIENTES COM GRAU 1 OU 2

Paciente	SVM Correlacionado	SVM	Categoria	Distância A	Distância B
1	0,0980	0,0842	+1	0,9020	0,9158
2	0,1630	0,1767	+1	0,8370	0,8233
3	0,4024	0,4424	+1	0,5976	0,5576
4	1,2040	1,0650	+1	0,2040	0,0650
5	0,0225	0,0916	+1	0,9775	0,9084
6	0,6268	0,2926	+1	0,3732	0,7074
7	0,6284	0,3929	+1	0,3716	0,6071
8	0,6915	0,6505	+1	0,3085	0,3495
9	0,7974	0,6852	+1	0,2026	0,3148
10	-0,9295	-0,6741	-1	0,0705	0,3259
11	-0,5912	-0,3690	-1	0,4088	0,6310
12	-0,2671	-0,2527	-1	0,7329	0,7473
13	-0,5666	-0,3585	-1	0,4334	0,6415
14	-0,5116	-0,4216	-1	0,4884	0,5784
15	-0,4824	-0,5305	-1	0,5176	0,4695
16	-1,3532	-1,3878	-1	0,3532	0,3878
17	-0,4016	-0,3029	-1	0,5984	0,6971
18	-0,1385	-0,0029	-1	0,8615	0,9971

FONTE: O autor (2016).

Os valores associados às Distâncias A e B da (TABELA 7) referem-se a diferença entre os valores obtidos pelo método SVM Correlacionado e SVM convencional e as classes corretas, respectivamente, de cada paciente utilizado na fase de teste dos métodos utilizados.

Já a classificação feita pelo método do discriminante de Fisher obteve classificação correta em 16 dos 18 padrões testados, se situando com um desempenho inferior aos demais métodos testados, evidenciando o bom desempenho do método SVM Correlacionado e demonstrando a superioridade dos resultados frente aos obtidos pelos outros dois métodos.

A (TABELA 8) apresenta um resumo dos resultados obtidos pelos métodos SVM Correlacionado, SVM e Discriminante Linear de Fisher, onde são expostos os parâmetros ótimos obtidos nos ajustes dos modelos, o erro *RMSE*, coeficiente de correlação *R* e o erro *MAPE* para os métodos.

TABELA 8 – RESUMO DOS PARÂMETROS E ERROS (GRAU 1 CONTRA 2)

Método	C	σ	$RMSE$	R	$MAPE$
SVM	4585,0	322,5	0,5730	0,8421	100,0%
SVM Correlacionado	1400,0	230,0	0,6577	0,8035	100,0%
Discriminante Linear de Fisher	-	-	-	-	86,36%

FONTE: O autor (2016).

5.1.3 Separação entre graus 3 e 4 pelo método SVM *Bootstrap*

Devido ao número reduzido de pacientes com os graus 3 e 4 de Ceratocone, utilizou-se o método SVM *Bootstrap*, uma vez que esse método foi desenvolvido basicamente para ser aplicado em abordagens em que o número de observações é reduzido.

Apenas quatro pacientes apresentaram grau 3 nos exames feitos na córnea e cinco pacientes foram classificados com grau 4 da doença Ceratocone, evidenciando o número baixo para poder efetuar a separação entre os conjuntos de treinamento, teste e validação. Para a aplicação do SVM convencional, separou-se duas amostras de pacientes de grau 3 e três de grau 4 para o treinamento, sobrando duas amostras de cada grau para formar o conjunto de testes.

Devido o número reduzido de amostras, ficou inviável a classificação pelos métodos SVM convencional e pelo Discriminante Linear de Fisher, sendo que partição do conjunto em dois menores resultaria em um número de cinco amostras para o ajuste e quatro para a fase de teste, sendo considerado conjuntos com numero insuficiente de amostra para efetuar as classificações de forma razoável. Devido a esse problema verifica-se que tanto a aplicação do SVM convencional como do Discriminante de Linear Fisher levam desvantagem em relação ao SVM *Bootstrap* por depender da separação dos dados conhecidos em dois grupos, problema que não ocorre no SVM *Bootstrap*, onde os dados conhecidos à priori são utilizados apenas para a fase de teste, aumentando o poder de generalização, como ficou evidenciado aqui.

Para a aplicação do LS-SVM *Bootstrap*, primeiramente deve-se destacar os máximos e mínimos de cada uma das cinco variáveis independentes, K central, Kisa (%), CV (%), Densidade e Hexanogalidade (%), para cada um dos padrões de

treinamento. Assim, convencionou-se a categoria +1 sendo os valores de grau 3 da doença, onde os resultados dos valores mínimos e máximos estão listados na (TABELA 9) dada a seguir.

TABELA 9 – MÍNIMO E MÁXIMO PARA CADA UMA DAS VARIÁVEIS INDEPENDENTES GRAU 3

Variável	Mínimo	Máximo	Amplitude
K central (D)	53,36	55,91	2,55
Kisa	43,24	119,99	76,75
CV	0,283	0,365	0,082
Densidade	2604,5	3025	420,5
Hexanogalidade	0,22	0,8	0,58

FONTE: O autor (2016).

Os valores de mínimo e máximo para as cinco características independentes para os pacientes classificados com grau 4 de Ceratocone estão listados na (TABELA 10) apresentados juntamente com a amplitude de cada intervalo.

TABELA 10 – MÍNIMO E MÁXIMO PARA CADA UMA DAS VARIÁVEIS INDEPENDENTES GRAU 4

Variável	Mínimo	Máximo	Amplitude
K central (D)	56,16	68,07	11,91
Kisa	8,22	2034,67	2026,45
CV	0,289	0,436	0,147
Densidade	2746,8	3823	1076,2
Hexanogalidade	0,49	0,69	0,2

FONTE: O autor (2016).

A maior vantagem da aplicação LS-SVM *Bootstrap* essa na fase de treinamento, uma vez que ela é feita com base nas reamostras *Bootstrap*, deixando os valores originais apenas para a fase de teste, uma vez que esse método foi desenvolvido principalmente para aplicações onde se possui um número reduzido de dados, que, dependendo da área estudada, é o cenário mais comum.

A (TABELA 11) apresenta os valores para as classificações feitas pelo método SVM *Bootstrap*, ao lado das classes corretas para cada paciente da fase de teste e a distância para o valor ideal.

TABELA 11 – VALORES OBTIDOS DAS CLASSIFICAÇÕES (1 CONTRA 2)

Paciente	Valor SVM <i>Bootstrap</i>	Categoria	Distância do ideal
1	1,1362	+1	0,1362
2	0,9368	+1	0,0632
3	0,8184	+1	0,1816
4	0,9363	+1	0,0637
5	-0,6835	-1	0,3165
6	-0,1851	-1	0,8149
7	-0,8513	-1	0,1487
8	-0,7571	-1	0,2429
9	-0,2245	-1	0,7755

FONTE: O autor (2016).

Ao todo foram criados $B = 1.000$ conjuntos de treinamento *Bootstrap* composto, número que minimizou o erro *RMSE* dentre todos os valores para B testados, de $n = 9$ amostras cada um, somando 9.000 amostras *Bootstrap* para a fase de treinamento e as nove amostras originais foram utilizadas na fase de teste. As constantes que retornaram os menores erros foram $C = 260,00$ e $\sigma = 166,9$ gerando o erro $RMSE = 0,4091$, com o erro ajustado $MAPE = 0,0\%$ e coeficiente de correlação $R = 0,9584$, classificando todos os nove padrões corretamente e demonstrando a eficiência do método proposto.

Em outro modelo ajustado utilizando $B = 10.000$ conjuntos de treinamento *Bootstrap*, gerando ao todo 90.000 amostras *Bootstrap*, o tempo de processamento, mesmo sendo baixo, aumentou consideravelmente, passando de 0,041 segundos para 0,4638 segundos, e o erro também acabou aumentando para $RMSE = 0,4622$, não demonstrando melhora nos resultados do modelo, e apenas um aumento no tempo de processamento.

Infelizmente, para este conjunto de dados não se foi possível criar uma comparação com resultados de outros métodos, devido ao número reduzido de dados, impossibilitando, ou dificultando, a partição do conjunto inicial nos subconjuntos de ajuste e teste.

5.2 RESERVATÓRIO DE MANASGAON

Inicialmente o objetivo era aplicar a metodologia desenvolvida a dados dos reservatórios do estado de São Paulo, por estarem passando por problemas de abastecimento e seca, porém, após cerca de um ano de tentativas em ter acesso aos dados sem sucesso, resolveu-se utilizar os dados do reservatório de Manasgaon, na Índia, coletados na literatura (DESWAL, PAL, 2008), ocorrendo à possibilidade de se comparar os resultados obtidos, criando assim mais uma forma de avaliação para a metodologia desenvolvida.

5.2.1 Previsão de evaporação de água pelo método SVR Correlacionado

Para esta aplicação, utilizou-se o método SVR Correlacionado, associado ao método das Margens Flexíveis *Bootstrap*, para modelos SVR, assim as margens são simétricas em relação às previsões efetuadas pelo SVR Correlacionado, apresentando então dois métodos abordados a esse conjunto de dados que trata o problema de se calcular previsões referentes a valores do nível de evaporação de água em reservatórios.

Ao se desenvolver a fase de treinamento, determinaram-se os melhores valores para os parâmetros C e σ , fixados em $C = 136,42$ e $\sigma = 18,453$, gerando o erro de $RMSE = 0,2792$, para a fase de teste e o coeficiente de correlação $R = 0,9242$. Já para a fase de treinamento, obteve-se o coeficiente de correlação $R = 0,9997$ e $RMSE = 0,087$, podendo ser comparado aos obtidos na literatura, na mesma fase de treinamento, $R = 0,960$ e $RMSE = 0,865$ (DESWAL, PAL, 2008) e $R = 0,996$ e $RMSE = 0,539$ (SAMUI, 2011), atendendo ao esperado, ou seja, a metodologia proposta se demonstrou eficiente e superior às demais já conhecidas e consolidadas.

Em relação ao tamanho B do conjunto *Bootstrap* de treinamento, esta aplicação obteve um tempo aproximado de 0,5811s, para $B = 1.000$ conjuntos *Bootstrap* para a fase de treinamento, e gastou 5,0544s para chegar à solução final quando utilizada uma reamostragem com 10.000 conjuntos *Bootstrap*. Porém, não houve redução significativa no erro cometido. A seguir, são apresentados os valores obtidos para a replicação $B = 1.000$ conjuntos.

A (TABELA 12) apresenta os valores das correlações das variáveis independentes e dos valores da evaporação que foram calculadas pelo método SVR Correlacionado.

TABELA 12 – CORRELAÇÕES ENTRE AS VARIÁVEIS (EVAPORAÇÃO)

	Temperatura	Velocidade do vento	Horas de sol	Umidade relativa do ar
Evaporação	0,8398	0,7162	0,3610	-0,5457

FONTE: O autor (2016).

A (TABELA 13) apresenta os valores reais coletados na literatura, valores previstos pelo método SVR convencional e as margens flexíveis, bem como as amplitudes calculadas entre as margens.

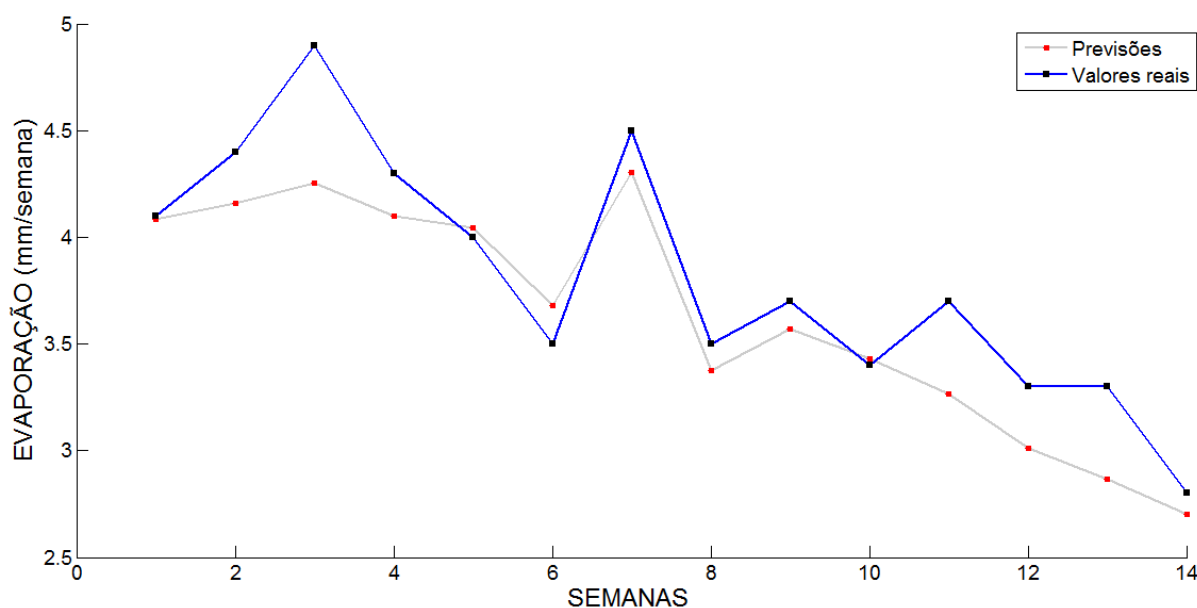
TABELA 13 – MÉTODO LS-SVR *BOOTSTRAP* PARA MARGENS FLEXÍVEIS

Semana	Valores reais	Margem inferior	Previsões	Margem superior	Amplitude	Erro padrão <i>Bootstrap</i>
1	4,1	3,39554	4,085185	4,83856	1,44302	0,75652
2	4,4	3,03026	4,161206	5,20734	2,17708	1,31294
3	4,9	3,16112	4,252201	5,09174	1,93062	1,04188
4	4,3	3,22640	4,100105	5,13385	1,90745	0,96341
5	4,0	3,31529	4,046527	5,09241	1,77712	0,73298
6	3,5	3,29381	3,681607	4,85518	1,56137	0,64041
7	4,5	3,55405	4,305958	4,98366	1,42961	0,55121
8	3,5	2,96179	3,375955	4,56116	1,59937	0,62718
9	3,7	3,22838	3,569742	4,09771	0,86933	0,42846
10	3,4	3,16716	3,428827	3,80797	0,64081	0,39887
11	3,7	2,84930	3,267068	3,36064	0,51134	0,37077
12	3,3	2,85150	3,010028	3,31729	0,46579	0,23561
13	3,3	2,69126	2,867451	3,13737	0,44611	0,24960
14	2,8	2,74082	2,702021	3,27005	0,52923	0,28374

FONTE: O autor (2016)

A (FIGURA 20) apresenta a comparação entre os valores obtidos pela aplicação do SVR correlacionado e os valores reais para as evaporações.

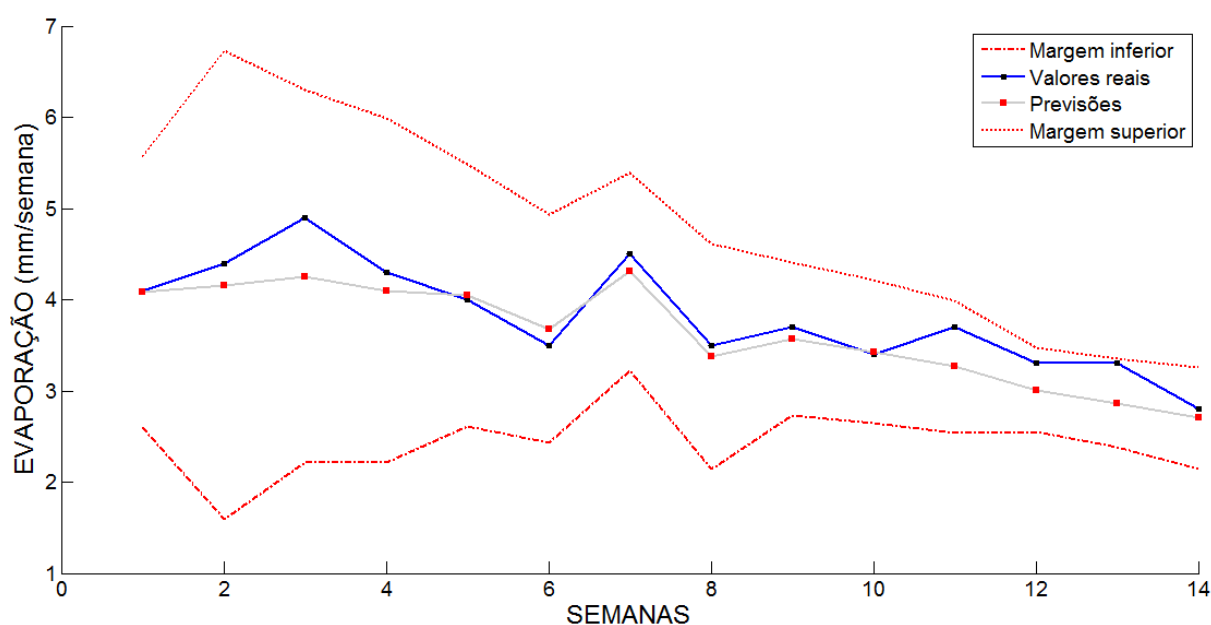
FIGURA 20 – COMPARAÇÃO GRÁFICA ENTRE PREVISÕES E VALORES REAIS



FONTE: O autor (2016).

Considerando as margens flexíveis *Bootstrap*, todos os valores reais se compreenderam dentro das margens, demonstrando o poder de generalização do método *Bootstrap*, e demonstrando a robustez do modelo.

A (FIGURA 21) apresenta a comparação entre os valores previstos, os valores reais e a introdução das margens flexíveis *Bootstrap*.

FIGURA 21 – MARGENS FLEXÍVEIS *BOOTSTRAP* PARA O MODELO LS-SVR

FONTE: O autor (2016).

A (TABELA 14) apresenta um resumo dos resultados obtidos pelo método SVR Correlacionado a fim de comparar com os resultados obtidos nos trabalhos de Deswal e Pal (2008) e Samui (2011), como forma de demonstração do desempenho do método proposto nesta tese.

TABELA 14 – RESUMO DOS PARÂMETROS E ERROS (EVAPORAÇÃO)

Método ou autor	C	σ	$RMSE$	R
SVR Correlacionado	136,42	18,453	0,27916	0,9242
LSSVR (SAMUI, 2011)	100,0	2,0	0,539	0,996
Rede Neural Artificial (DESWAL, PAL, 2008)	-	-	0,865	0,960

FONTE: O autor (2016).

5.3 ASTIGMATISMO

Foram abordados 26 pacientes com certo grau de astigmatismo em um dos olhos. Para esta aplicação utilizou-se o método SVR Correlacionado, e a abordagem das margens flexíveis *Bootstrap*. Além da aplicação, os dados foram confrontados com os valores calculados pelo método SVR convencional. Foram separadas 20, das 26, observações para a fase de treinamento, deixando as 6 restantes para a fase de teste.

5.3.1 Previsão do grau de astigmatismo pelo método SVR *Bootstrap*

Os parâmetros C e σ que retornaram os menores valores para o erro $RMSE$ foram $C = 29,7$ e $\sigma = 2,2$, gerando o erro igual à $RMSE = 0,04951$, para a aplicação que conta com a multiplicação pelo coeficiente de correlação. Já para a abordagem convencional, os valores ótimos para os parâmetros C e σ ficaram bem próximos aos do SVR Correlacionado, sendo $C = 39,01$ e $\sigma = 5,09$, porém o valor do erro cresceu para $RMSE = 0,07350$, se mantendo baixo, porém, maior que o valor alcançado pelo SVR correlacionado.

A (TABELA 15) a seguir apresenta os valores de correlação entre as variáveis quantitativas das características Ceratometria e Refração com os valores do Astigmatismo examinados.

TABELA 15 – CORRELAÇÕES ENTRE VARIÁVEIS (ASTIGMATISMO)

	Ceratometria		Refração	
	Mais plano	Mais curvo	Esfera	Cilíndrico
Astigmatismo	-0,7780	0,4481	0,4953	-0,8066

FONTE: O autor (2016).

A (TABELA 16) apresenta os valores para as previsões feitas pelo método SVR, SVR Correlacionado e os valores reais para o grau de astigmatismo dos dados utilizados, bem como as distâncias dos resultados do método SVR aos valores reais (Distância A) e do método SVR Correlacionado (Distância B).

TABELA 16 – COMPARAÇÃO ENTRE OS MÉTODOS SVR E SVR CORRELACIONADO

SVR	SVR Correlacionado	Valores reais	Distância A	Distância B
1,6805	1,6659	1,75	0,0695	0,0841
0,6428	0,7232	0,75	0,1072	0,0268
0,9766	1,0062	1,00	0,0234	0,0062
0,5525	0,5809	0,50	0,0525	0,0809
1,0165	1,0108	1,00	0,0165	0,0108
1,3619	1,2646	1,25	0,1119	0,0146

FONTE: O autor (2016).

A (TABELA 17) apresenta os valores do SVR Correlacionado, os valores dos erros *Bootstrap*, das margens *Bootstrap* para as regressões calculadas e os valores reais. Os valores marcados com um asterisco (*) foram modificados devido o valor da margem inferior ser negativa, e o valor do astigmatismo ser sempre positivo.

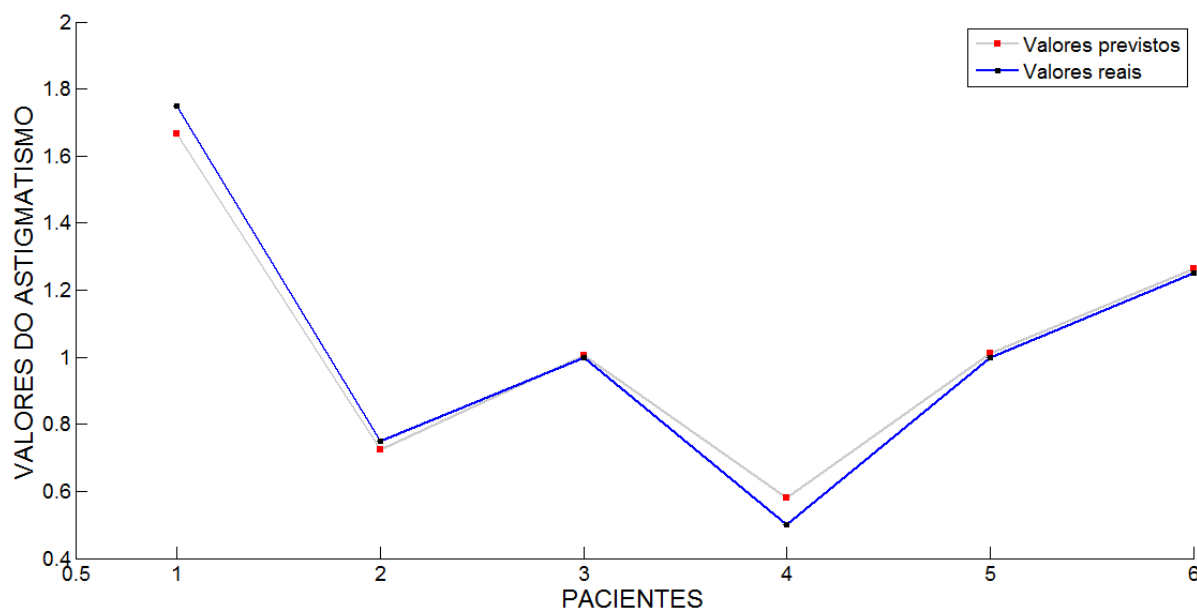
TABELA 17 – MARGENS FLEXÍVEIS *BOOTSTRAP* PARA O SVR

SVR correlacionado	Erro padrão <i>bootstrap</i>	Valores reais	Margem inferior	Margem superior
1,6659	0,9101	1,75	-0,1189* (0)	3,437765
0,7232	0,6467	0,75	-0,5402* (0)	1,977764
1,0062	0,6044	1,00	-0,1643* (0)	2,189301
0,5809	1,0162	0,50	-1,4152* (0)	2,569656
1,0108	0,5542	1,00	-0,0690* (0)	2,099629
1,2646	0,6918	1,25	-0,0871* (0)	2,592181

FONTE: O autor (2016).

A (FIGURA 22) apresenta as projeções dos valores calculados, bem como os valores reais para os exames feitos referentes ao grau de astigmatismo para os seis pacientes que fazem parte do conjunto de teste.

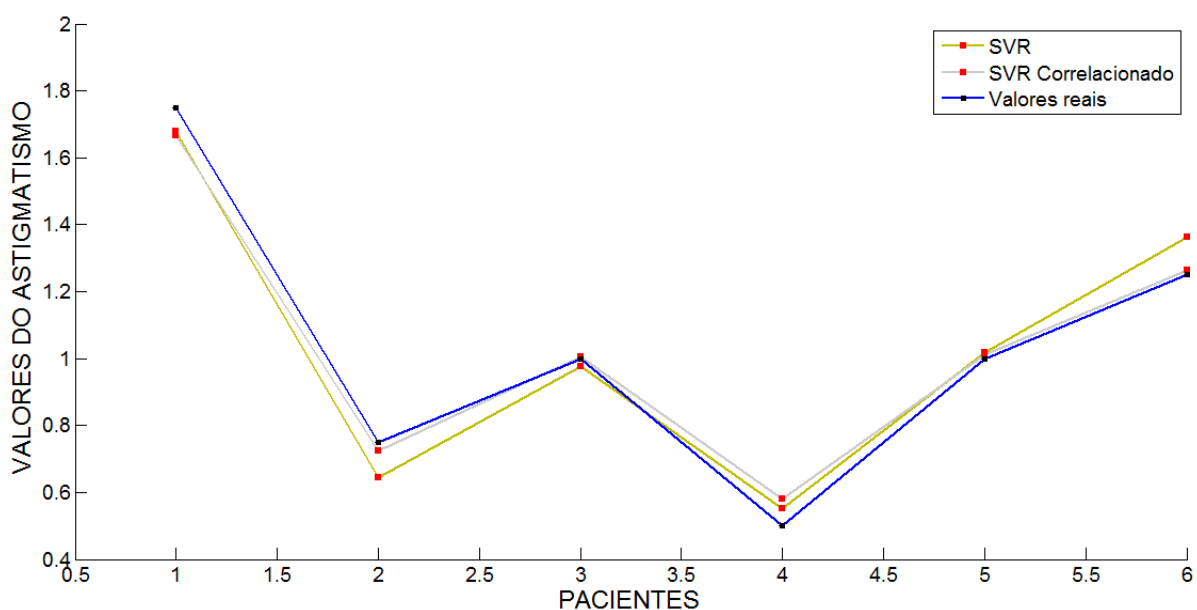
FIGURA 22 – PREVISÕES DO SVR CORRELACIONADO PARA OS SEIS PACIENTES



FONTE: O autor (2016).

A (FIGURA 23) apresenta os valores do método SVR, SVR Correlacionado e os valores reais para os mesmo seis pacientes.

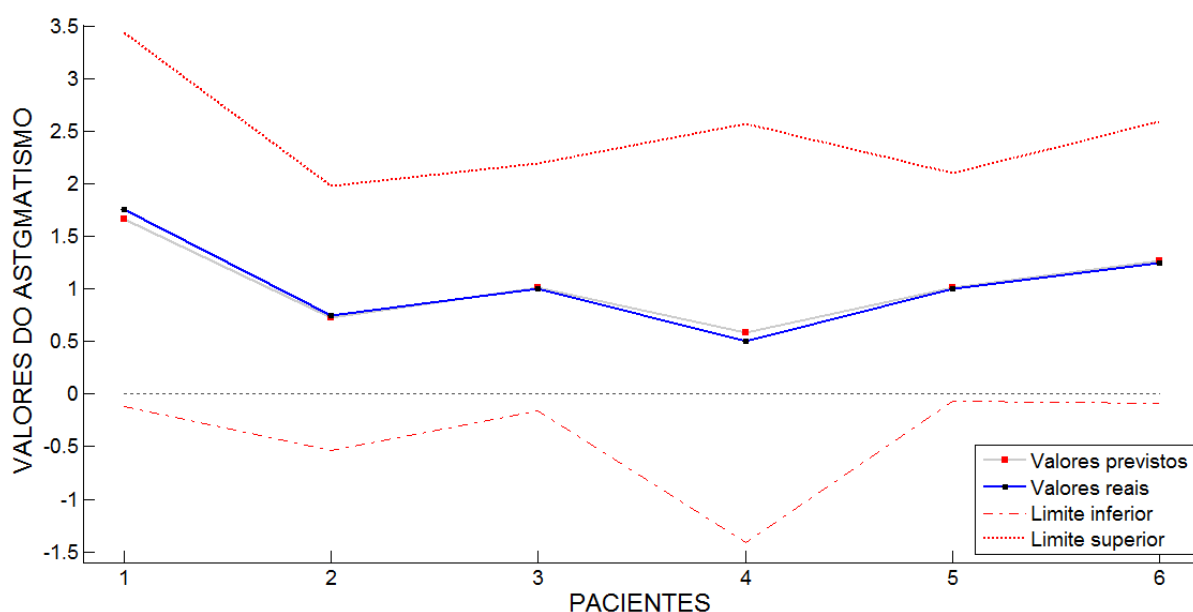
FIGURA 23 – PREVISÕES E VALORES REAIS PARA OS SEIS PACIENTES



FONTE: O autor (2016).

A (FIGURA 24) apresenta as margens flexíveis *Bootstrap* para a aplicação dos dados da previsão do astigmatismo. A linha tracejada paralela ao eixo dos pacientes mostra o real limite das margens flexíveis, uma vez que os valores do astigmatismo não podem ser negativos, devido ao seu eixo que varia de 0 a 180 graus.

FIGURA 24 – MARGENS FLEXÍVEIS *BOOTSTRAP* PARA AS PREVISÕES



FONTE: O autor (2016).

A (TABELA 18) apresenta o resumo dos valores ótimos para os parâmetros e os resultados dos erros obtidos pelo método SVR Correlacionado e SVR convencional, a fim de comparar com os resultados obtidos e verificar o desempenho do método proposto pelo trabalho na aplicação para a previsão do grau de astigmatismo dos seis pacientes do grupo de teste.

TABELA 18 – RESUMO DOS PARÂMETROS E ERROS (ASTIGMATISMO)

Método ou autor	C	σ	$RMSE$	R
SVR Correlacionado	136,42	18,453	0,04951	0,9965
SVR	39,01	5,09	0,07350	0,9824

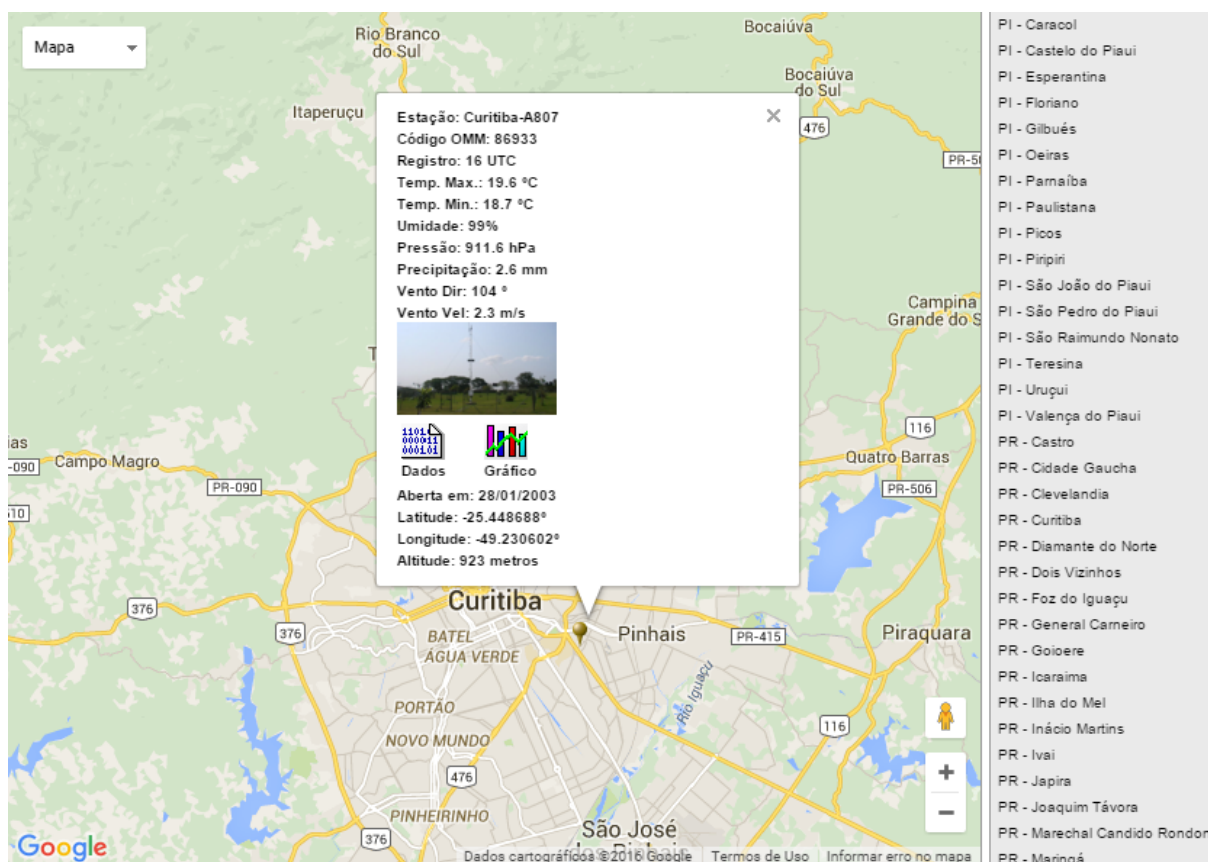
FONTE: O autor (2016).

5.4 PREVISÃO DA TEMPERATURA

Para os valores observados a partir das estações meteorológicas de superfície automática, o sítio eletrônico do INMET libera apenas os valores dos últimos 90 dias para *download*. Assim, coletou-se os valores de 84 dias seguidos, compreendidos entre o dia 25 de outubro de 2015 e 16 de janeiro de 2016, separados por semanas, e utilizados os valores médios das semanas para os valores de umidade relativa (%), temperatura do ponto de orvalho ($^{\circ}$ C), pressão atmosférica (hPa), velocidade do vento (m/s) e rajada (m/s), e radiação solar (kJ/m^2). Já para os valores da precipitação, utilizou-se o acumulado para a semana.

A (FIGURA 25) apresenta a região geográfica da coleta dos dados aplicados ao método SVR *Bootstrap*, na região de Curitiba, a partir da Estação Meteorológica de Observação de Superfície Automática, situada no Centro Politécnico da Universidade Federal do Paraná.

FIGURA 25 – REGIÃO DA ESTAÇÃO DE COLETA DOS DADOS



FONTE: Adaptado de INMET (2016).

Para calcular as previsões dos valores da temperatura em ° C utilizou-se o método SVR *Bootstrap*, uma vez que tal método cria a possibilidade de calcular as previsões para as doze semanas coletadas, usando os dados coletados apenas na fase de teste.

Também foi aplicado o método SVR correlacionado, porém, por depender dos dados já a partir da fase de treinamento, foi necessário criar uma divisão no conjunto formado pelas 12 médias semanais utilizando 7 semanas para a fase de treinamento e as 5 restantes para a fase de teste.

As correlações referentes entre as variáveis estão apresentadas na (TABELA 19), onde foram utilizadas apenas no método SVR correlacionado, uma vez que não existe a etapa da multiplicação pela correlação no método SVR *Bootstrap*.

TABELA 19 – CORRELAÇÕES ENTRE AS VARIÁVEIS (TEMPERATURA)

	Umidade	Ponto de orvalho	Pressão	Velocidade do vento	Rajada	Radiação solar	Chuva
T (° C)	-0,8188	0,8228	-0,2800	0,2619	0,3831	0,8114	0,3048

FONTE: O autor (2016).

Para o método SVR Correlacionado, utilizaram-se os seguintes valores para os parâmetros de treino: $C = 115,0$ e $\sigma = 495,0$. O método retornou um valor para o erro de $RMSE = 1,6647$. Para as cinco previsões feitas no conjunto de teste, o erro para o SVR convencional ficou em $RMSE = 1,9231$.

Já para o método SVR *Bootstrap*, a calibração dos parâmetros fixou os melhores resultados para $B = 510$, $C = 23,0$ e $\sigma = 116,0$, gerando o seguinte valor para o erro $RMSE = 1,2229$, sendo que para esse modelo, utilizaram-se as 12 semanas para a fase de teste, uma vez que na fase de treinamento o método utiliza apenas os valores máximos e mínimos para cada característica para gerar os conjuntos *Bootstrap* de treinamento.

Os limites para as sete características que foram utilizadas para a fase de treinamento do modelo estão dispostas na (TABELA 20) a seguir, bem como a amplitude dos intervalos onde foram feitos os sorteios *Bootstrap*.

TABELA 20 – CARACTERÍSTICAS PARA O TREINAMENTO *BOOTSTRAP*

Característica	Limite inferior	Limite superior	Amplitude
Temperatura	16,7884	22,4887	5,7003
Umidade	0,7146	0,9014	0,1868
Ponto de orvalho	15,0905	19,0384	3,9479
Pressão	907,1830	913,2833	6,1003
Velocidade vento	1,6232	2,4976	0,8744
Rajada	4,5042	6,0988	1,5946
Radiação solar	411,0296	975,2732	564,2436
Chuva	17,8	103,2	85,4

FONTE: O autor (2016)

A (TABELA 21) apresenta os valores das previsões calculadas, bem como o valor real para a temperatura, os erros *Bootstrap* e os intervalos das margens flexíveis *Bootstrap*.

TABELA 21 – MARGENS FLEXÍVEIS *BOOTSTRAP* PARA O SVR

SVR <i>Bootstrap</i>	Erro padrão <i>Bootstrap</i>	Valores reais	Margem inferior	Margem superior
19,2401	1,0217	17,9262	17,2376	21,2426
18,8854	1,1771	16,7884	16,5783	21,1926
20,8438	1,0163	20,8330	18,8518	22,8357
20,0276	1,0109	19,8369	18,0462	22,0090
18,7150	1,0185	18,7955	16,7187	20,7113
19,9685	1,0273	18,4458	17,9550	21,9820
19,2372	1,0230	20,3875	17,2322	21,2423
20,8503	1,0083	22,4887	18,8740	22,8265
21,3618	1,0119	22,2893	19,3785	23,3450
21,5576	1,0117	22,0179	19,5747	23,5404
19,5821	1,0227	21,1074	17,5777	21,5865
20,1073	1,0254	21,5409	18,0975	22,1172

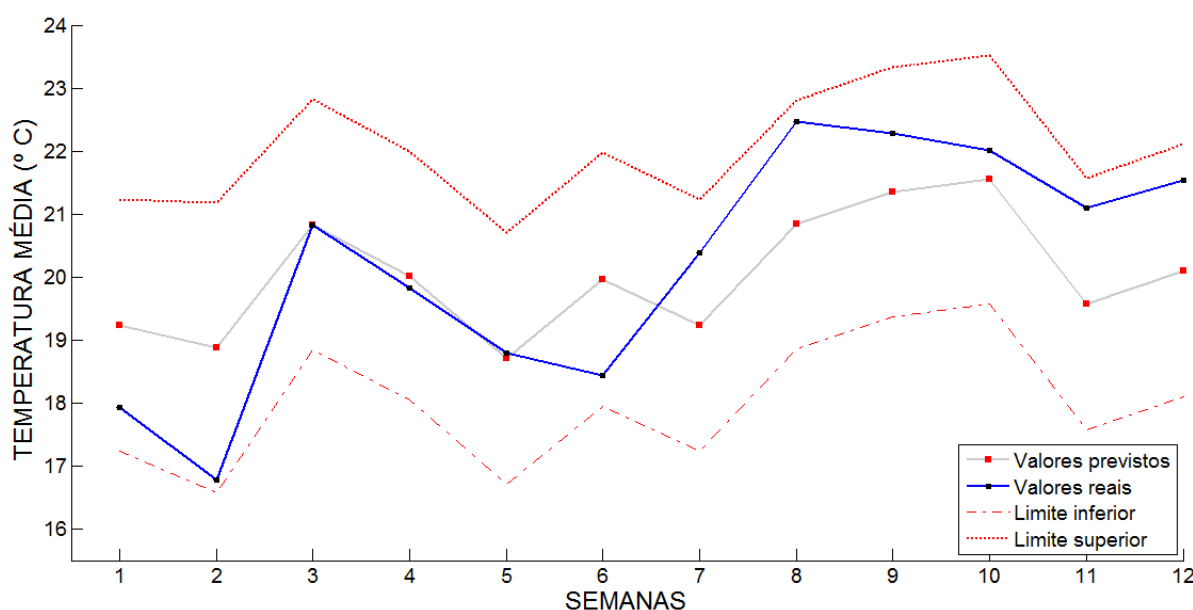
FONTE: O autor (2016)

Fazendo a comparação entre o método SVR *Bootstrap* e o SVR Correlacionado para as últimas cinco previsões para as semanas, uma vez que o método correlacionado permitiu apenas a previsão de cinco semanas na fase de teste, devido a sua partição nos conjuntos de dados de ajuste e teste, obteve-se o erro $RMSE = 1,2758$ para as cinco últimas previsões calculadas pelo SVR *Bootstrap*, contra $RMSE = 1,6647$ do método correlacionado, apresentando uma melhora

significativa nas previsões. Para o método SVR convencional, o erro ficou acima dos calculados pelos métodos desenvolvidos, $RMSE = 1,7503$.

A (FIGURA 26) apresenta graficamente as previsões e as novas margens flexíveis *Bootstrap* para as previsões calculadas pelo método SVR *Bootstrap*. É possível visualizar a eficiência do método, uma vez que todos os valores reais se situam dentro das margens *Bootstrap*, resultado ideal esperado.

FIGURA 26 – PREVISÃO E MARGENS *BOOTSTRAP* DA TEMPERATURA



FONTE: O autor (2016).

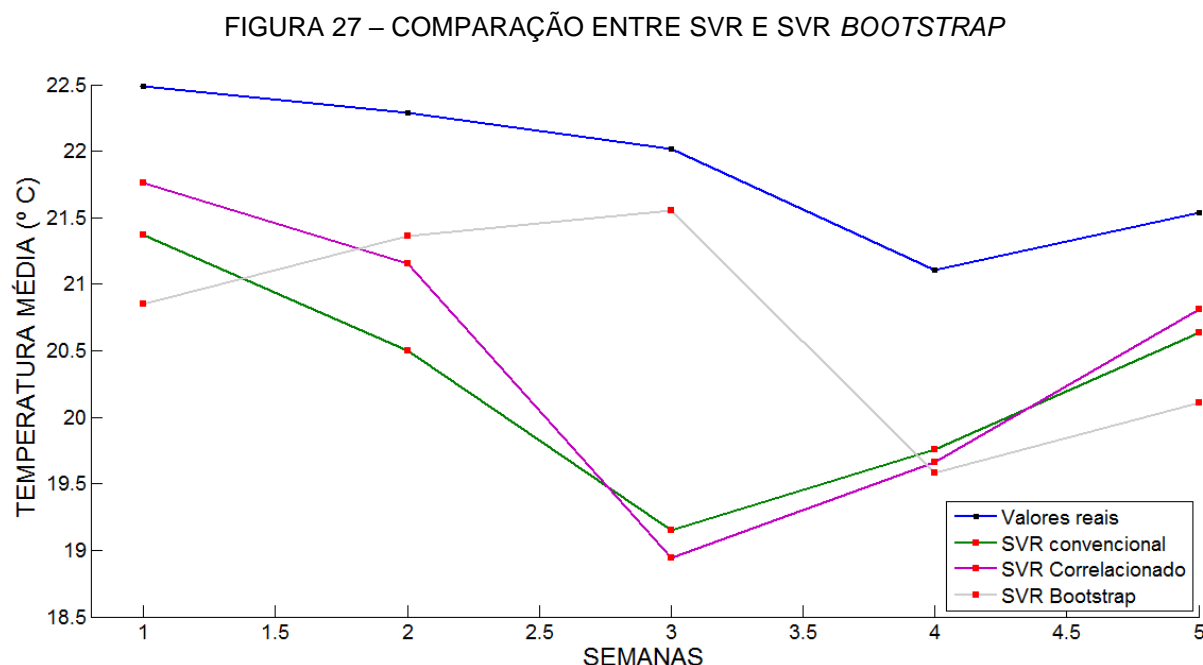
A (TABELA 22) apresenta as previsões feitas pelos métodos SVR convencional, SVR Correlacionado e SVR *Bootstrap*, e os valores reais observados para as cinco últimas semanas da fase de teste.

TABELA 22 – VALORES CALCULADOS PARA AS 5 PREVISÕES SEMANAIS

Valores reais	SVR	SVR Correlacionado	SVR <i>Bootstrap</i>
22,4887	21,3727	21,7597	20,8503
22,2893	20,5012	21,1563	21,3618
22,0179	19,1474	18,9471	21,5576
21,1074	19,7597	19,6654	19,5821
21,5409	20,6362	20,8119	20,1073

FONTE: O autor (2016)

A (FIGURA 27) apresenta os valores das previsões e os valores reais da temperatura para as cinco semanas, calculadas pelo método SVR convencional, SVR Correlacionado e SVR *Bootstrap*.



FONTE: O autor (2016).

Com base na (TABELA 22) e (FIGURA 27), fica evidente o poder do método SVR *Bootstrap*, até mesmo quando comparado ao método SVR correlacionado, que por sua vez já retornou valores melhores que os obtidos pelo método SVR convencional, o seu erro foi o menor. Assim, é razoável considerar o método SVR *Bootstrap* robusto e com alto grau de generalização, obtendo-se resultados satisfatórios na aplicação para a previsão da temperatura, com base nos dados coletados para a cidade de Curitiba, entre outubro de 2015 e janeiro de 2016.

A (TABELA 23) apresenta o resumo dos valores ótimos para os parâmetros e os resultados dos erros obtidos pelos métodos SVR *Bootstrap*, SVR Correlacionado e SVR convencional, para as ultimas cinco semanas, intervalo de tempo onde se fez possível à comparação entre os métodos, a fim de verificar o desempenho do método proposto pelo trabalho na aplicação para a previsão da média da temperatura semanal.

TABELA 23 – RESUMO DOS PARÂMETROS E ERROS (TEMPERATURA)

Método ou autor	C	σ	$RMSE$	R
SVR	36,2	45,3	1,7503	0,4272
SVR Correlacionado	115,0	495,0	1,6647	0,5293
SVR <i>Bootstrap</i>	23,0	116,0	1,2758	0,8206

FONTE: O autor (2016).

5.5 COTAÇÃO DO DÓLAR

Como já mencionado, nesta aplicação utilizou-se o método SVR *Bootstrap*, uma vez que o número de semanas abordadas somou 17 observações. Assim, o ajuste do modelo foi feito em cima das amostras *Bootstrap*, ficando as 17 amostras reais para a fase de teste do modelo, dos multiplicadores de Lagrange e seus parâmetros. Este período de tempo foi selecionado devido à maior suavidade nas variações da cotação.

Foram destacados, para o sorteio aleatório *Bootstrap* os valores máximos e mínimos das cotações de segunda a quinta e das variações. Tais valores são apresentados na (TABELA 24), bem como os valores máximos e mínimos para a variável dependente y .

TABELA 24 – CARACTERÍSTICAS PARA O TREINAMENTO *BOOTSTRAP* (DÓLAR)

Característica	Limite inferior	Limite superior	Amplitude
y	3,6960	4,1080	0,4120
$x_{segunda}$	3,7328	4,0487	0,3159
$x_{terça}$	3,7015	4,0680	0,3665
x_{quarta}	3,7365	4,1025	0,3660
x_{quinta}	3,7265	4,1644	0,4379
$x_{média}$	-0,0459	0,0442	0,0901

FONTE: O autor (2016)

Os parâmetros utilizados para o ajuste e teste foram $C = 1750,5$, $\sigma = 9,9899$ e $B = 1000$ replicações. O método retornou um valor para o erro de $RMSE = 0,10886$, e os valores das previsões estão listados na (TABELA 25), bem como os valores reais para os valores de y .

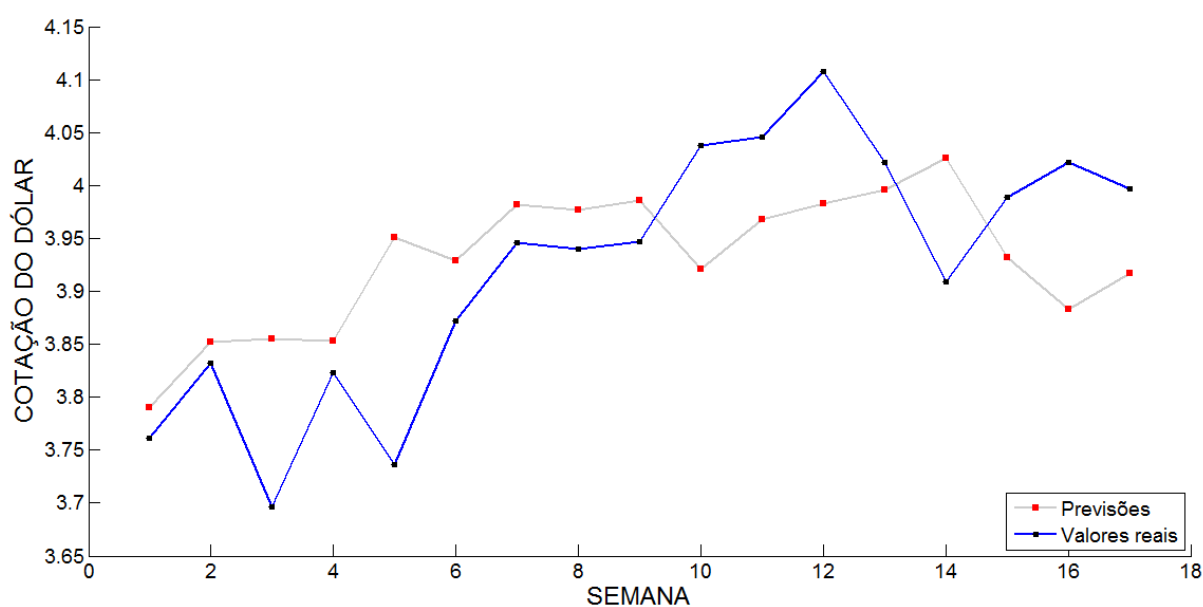
TABELA 25 – PREVISÕES PARA A COTAÇÃO DO DÓLAR

SVR <i>Bootstrap</i>	Valores reais	Distância do ideal
3,7907	3,7613	0,0294
3,8521	3,8324	0,0197
3,8554	3,6960	0,1594
3,8532	3,8232	0,0300
3,9506	3,7360	0,2146
3,9285	3,8721	0,0564
3,9816	3,9456	0,036
3,9769	3,9401	0,0368
3,9859	3,9470	0,0389
3,9209	4,0376	0,1167
3,9681	4,0453	0,0772
3,9831	4,1080	0,1249
3,9962	4,0217	0,0255
4,0254	3,9093	0,1161
3,9318	3,9889	0,0571
3,8828	4,0217	0,1389
3,9174	3,9972	0,0798

FONTE: O autor (2016)

A (FIGURA 28) apresenta a dispersão dos valores reais e das previsões feitas pelo método SVR *Bootstrap*. Nesta aplicação não se utilizou a aplicação do multiplicador da correlação.

FIGURA 28 – PREVISÕES DA COTAÇÃO DO DÓLAR



FONTE: O autor (2016).

A (TABELA 26) apresenta os valores para as previsões calculadas pelo método SVR *Bootstrap*, o erro padrão *Bootstrap*, os valores reais e as margens flexíveis *Bootstrap* para cada um dos valores previstos pelo método SVR *Bootstrap*. Para uma primeira análise, pode-se concluir que as margens *Bootstrap* abrangem todos os valores reais para a cotação do dólar, demonstrando assim a eficiência da aplicação das novas margens para modelos SVR, conseguindo situar 100% dos dados de teste dentro da sua amplitude.

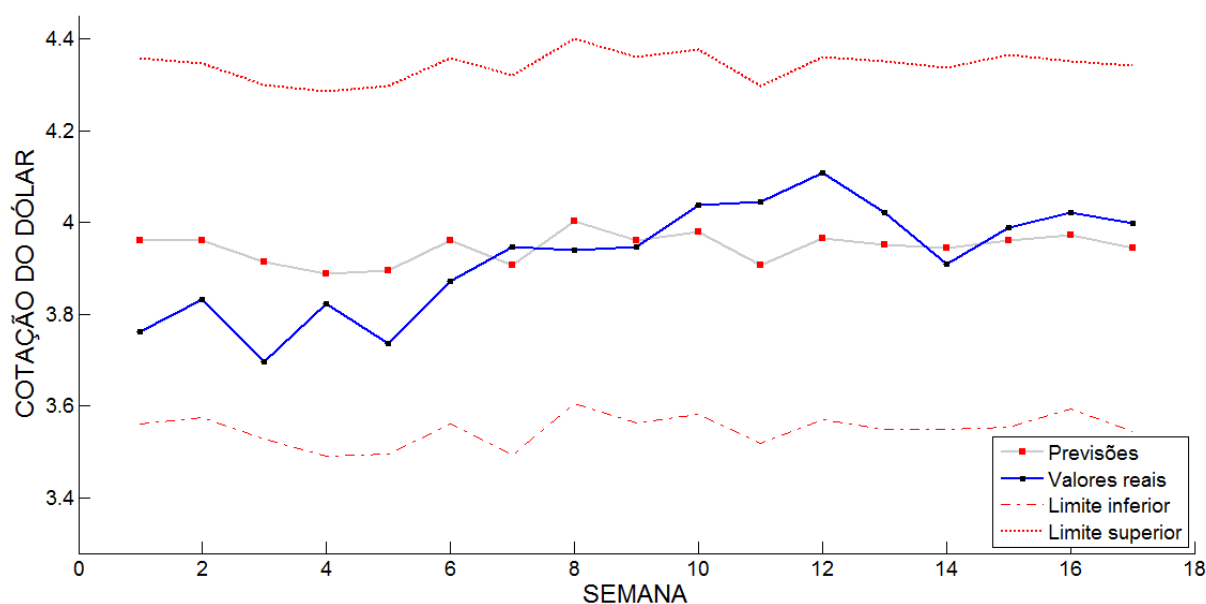
TABELA 26 – MARGENS FLEXÍVEIS PARA A COTAÇÃO DO DÓLAR

SVR <i>Bootstrap</i>	Erro padrão <i>Bootstrap</i>	Valores reais	Limite inferior	Limite superior
3,9599	0,2031	3,7613	3,5617	4,3581
3,9603	0,1967	3,8324	3,5748	4,3458
3,9148	0,1966	3,6960	3,5294	4,3002
3,8883	0,2030	3,8232	3,4903	4,2863
3,8955	0,2043	3,7360	3,4951	4,2959
3,9598	0,2032	3,8721	3,5615	4,3581
3,9071	0,2106	3,9456	3,4943	4,3199
4,0024	0,2026	3,9401	3,6053	4,3995
3,9618	0,2027	3,9470	3,5645	4,3591
3,9797	0,2030	4,0376	3,5819	4,3775
3,9076	0,1984	4,0453	3,5187	4,2965
3,9656	0,2013	4,1080	3,5712	4,3600
3,9506	0,2044	4,0217	3,5500	4,3512
3,9435	0,2005	3,9093	3,5505	4,3365
3,9605	0,2068	3,9889	3,5551	4,3659
3,9731	0,1929	4,0217	3,5951	4,3511
3,9435	0,2028	3,9972	3,5460	4,3410

FONTE: O autor (2016)

A (FIGURA 29) apresenta a versão gráfica da (TABELA 26), a fim de exemplificar os resultados obtidos pelo método SVR *Bootstrap* e as margens flexíveis *Bootstrap*. São apresentados os valores das previsões, valores reais e as margens flexíveis *Bootstrap* para o modelo.

FIGURA 29 – MARGENS FLEXÍVEIS PARA COTAÇÃO DO DÓLAR



FONTE: O autor (2016).

Assim, tem-se a aplicação na previsão da cotação do dólar no fechamento da sexta-feira das 17 semanas compreendidas entre os dias 03 de novembro de 2015 e 26 de fevereiro de 2016, mostrando também a robustez das margens flexíveis *Bootstrap*, abrangendo 100% dos valores reais dentro das 17 semanas, gerando nenhum erro referente às previsões estarem fora das margens flexíveis.

A (TABELA 27) apresenta um resumo a respeito dos parâmetros ótimos e os resultados obtidos, tais como o erro *RMSE* e o coeficiente de correlação *R*, obtidos para o método SVR *Bootstrap*.

TABELA 27 – RESUMO DOS PARÂMETROS E ERROS (DÓLAR)

Método	C	σ	$RMSE$	R
SVR <i>Bootstrap</i>	148,5	9,99	0,0968	0,5490

FONTE: O autor (2016).

6 CONCLUSÃO

O método *Support Vector Machine* é considerado uma técnica robusta por si só, e tanto utilizado para classificações, como para regressões, como mostrado nas aplicações apresentadas nesta tese.

Ao todo foram desenvolvidas três metodologias inovadoras: *SVM Bootstrap*, *SVM Correlacionado* e as margens *Bootstrap* para modelos SVR. As metodologias foram aplicadas para problemas de classificações de padrões e problemas de regressão não linear múltipla. Ao todo foram aplicados a cinco tipos de problemas distintos, a fim de verificar o poder de generalização.

Os modelos *SVM Bootstrap* foram diretamente aplicados aos conjuntos de dados onde o número de amostras era reduzido, utilizando o *Bootstrap* na fase de treinamento, gerando B novos conjuntos *bootstrap* de dados sintéticos, porém, que se compreendem entre os valores máximos e mínimos das amostras originais. Tal abordagem se demonstra mais eficiente que o treinamento normal dos modelos SVM por criar a possibilidade de se testar o modelo com todos os dados conhecidos, excluindo o problema da necessidade de dividir o conjunto de dados, que geralmente é reduzido, em dois ou mais conjuntos para treinamento, teste e validação.

A eficiência das metodologias ficou comprovada ao se comparar os resultados obtidos com os resultados de outros métodos, tais como os métodos SVM e SVR convencionais e o discriminante linear de Fisher, sempre retornando um erro menor que tais técnicas.

Optou-se por apresentar aplicações com resultados coletados na literatura principalmente para comparar os resultados e constatar a eficiência dos métodos desenvolvidos nesta tese.

Todas as soluções dos métodos SVM foram implementadas utilizando o método *Least Squares Support Vector Machines*, devido sua baixa complexidade matemática e computacional, frente aos problemas de otimização quadrática dos modelos SVM convencionais, e também a facilidade em ser implementada em linguagem Fortran, linguagem computacional utilizada em todas as modelagens apresentados neste trabalho.

Uma das maiores dificuldades, após as implementações dos métodos desenvolvidos, na abordagem dos problemas de classificação e regressão, pelos

modelos SVM, foi à calibração dos parâmetros C e σ para o modelo SVM correlacionado, uma vez que esta calibração deve ser feita pelo usuário, podendo ser automatizado; porém ainda existe o problema em determinar as melhores configurações na fase de treinamento, e para os modelos SVM *Bootstrap*, também foi necessário calibrar o valor da constante B , referente ao número de novos conjuntos *Bootstrap* para o modelo.

Todas as modelagens feitas nesta tese abordaram problemas ou dados reais, porém os dados utilizados na modelagem das previsões das perdas de água por evaporação em reservatórios foram coletados na literatura referenciada, e não pertencem a reservatórios brasileiros.

Durante os dois últimos anos do desenvolvimento desta tese, foram feitas pelo menos três tentativas em se obter dados dos reservatórios nacionais, duas em São Paulo e uma no Paraná, porém, devido a problemas burocráticos, as requisições não foram aceitas, ou os dados recebidos estavam incompletos ou, ainda, eram totalmente diferentes dos dados requisitados.

Ficou evidenciado os resultados positivos gerados pelos métodos desenvolvidos, com base nas aplicações e nas comparações com métodos já conhecidos e amplamente divulgados. Em todas as aplicações, os métodos desenvolvidos foram confrontados com pelo menos um método alternativo, e em excluindo os raros casos, que os resultados foram no mínimo equivalentes, os resultados foram melhores quando confrontados aos resultados dos métodos já consagrados.

A principal vantagem observada para o método SVM Correlacionado foi à facilidade em se desenvolver o método, uma vez que a sua alteração, anterior à fase de treinamento, consiste em criar uma nova constante reguladora para os dados de entrada, e embora pareça uma modificação elementar, retornou resultados melhores que os métodos convencionais, que não possuem tal fase pré-treinamento.

Ficou constatado, empiricamente, que a introdução das constantes de correlação não altera o valor absoluto da correlação entre as variáveis, apenas torna o valor da correlação, caso ela seja negativa, em positiva, e se a correlação já for positiva o valor não se altera. Além disso, em nenhuma aplicação, os resultados foram inferiores aos calculados pelos métodos SVM convencionais, demonstrando assim a eficiência da metodologia proposta.

Para o método SVM *Bootstrap*, pode-se destacar a vantagem de se poder trabalhar com o número de amostras que for necessário, devido ao processo de reamostragem *Bootstrap*. Tal condição foi aplicada ao se deparar com problemas com número reduzido de dados, uma vez que a análise acaba sendo comprometida pelo número reduzido de dados.

Assim, desenvolveu-se a modificação para a fase de treinamento, baseando no processo de reamostragem *Bootstrap*, porém, aumentando o conjunto de amostras distintas, uma vez que o processo comum do *Bootstrap* utiliza apenas os valores conhecidos das amostras, já o desenvolvido cria um intervalo de variação para os sorteios *Bootstrap*. Também pode-se destacar a vantagem em se poder utilizar todos os dados originais para a fase de teste, onde melhora-se os resultados das calibrações dos parâmetros C e σ , deixando o modelo mais maleável e mais generalista.

Vale ressaltar também o poder do método SVR *Bootstrap*, que confrontado com o método SVR Correlacionado e o método SVR convencional obteve valores para o *RMSE* menores que para os dois métodos citados.

Para a aplicação na previsão da cotação do dólar, o método SVR *Bootstrap* obteve um erro $RMSE = 0,0968$, relativamente baixo, ficando evidente a robustez do método para as 17 semanas seguidas, além da robustez do novo método de obtenção de margens flexíveis para os modelos SVR. Pode-se levar em consideração que as observações foram coletadas no final do ano de 2015, onde ocorreram vários feriados e recessos, podendo ter prejudicado as previsões, porém vale ressaltar a eficácia da aplicação.

Para os próximos trabalhos, estuda-se aplicar os métodos desenvolvidos para os dados dos reservatórios de água nacionais, principalmente para os reservatórios do estado de São Paulo, devido aos problemas que o estado sofreu no ano de 2015, apresentando mais uma forma de monitoramento para os reservatórios.

Além disso, pretende-se aplicar o método de classificação de padrões a novas doenças, a fim de auxiliar o processo de prognósticos médicos diminuindo os erros e minimizando o tempo gasto com classificações erradas, e, quem sabe, ajudar a salvar vidas.

REFERÊNCIAS

- ABREU, A. L. E.; CHAVES NETO, A. Modelo para previsão de evaporação em reservatórios de água. **Anais do Simpósio de Métodos Numéricos Computacionais (SMNC)**, UFPR, 2015.
- ALMEIDA, E. D. **Classificação Ordinal com Opção de Rejeição**. Dissertação de Mestrado (Engenharia Informática e Computação). Faculdade de Engenharia da Universidade do Porto, Porto, PT, 2010.
- BORIN, A. **Aplicações de Máquinas de vetores de suporte por mínimos quadrados (LS-SVM) na quantificação de parâmetros de qualidade de matrizes lácteas**. Tese de Doutorado (Química). Universidade Estadual de Campinas, 2007.
- BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A Training Algorithm for Optimal Margin Classifiers. **COLT '92 Proceedings of the fifth annual workshop on Computational learning theory**, pag. 144-152. ACM New York, NY, USA. 1992.
- BOYD, Samuel. **Distrofias e degenerações corneais**. Ed. Jaypee-Highlights. Panama/República Dominicana. 2012.
- BURGES, C. J.C. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery. **Kluwer Academic Publishers, Boston. Manufactured in The Netherlands**, 2, 121–167, 1998.
- BUZZI, M. F. **Uma nova proposta para o treinamento não supervisionado em redes neurais de base radial para previsão de séries temporais**, Tese de Doutorado (Métodos Numéricos em Engenharia). Universidade Federal do Paraná, 2012.
- CAMILO, C. O. & SILVA, J. C. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. Relatório Técnico, Instituto de Informática Universidade Federal de Goiás, 2009.
- CARVALHO, B. P. R. **Novas Estratégias para Detecção Automática de Vetores de Suporte em *Least Squares Support Vector Machines***. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Minas Gerais, MG, 2005.
- CARVALHO, B. P. R. O Estado da Arte em Métodos de Reconhecimento de Padrões: Support Vector Machines. **Congresso Nacional de Tecnologia e Comunicação**, 2005, Belo Horizonte, Minas Gerais.
- CHAVES NETO, A., **“Bootstrap” em Séries Temporais**. (Tese de Doutorado em Engenharia Elétrica). Pontifícia Universidade Católica do Rio de Janeiro. PUC-RJ, 1991.
- CHAVES NETO, A. **Análise de séries temporais (notas de aula)**. Departamento de Estatística, Universidade Federal do Paraná, Curitiba, 2009.

CORTES, C. VAPNIK, V. **Support-Vector Networks. Machine Learning**, Volume 20, Número 3, Setembro/1995, pag. 273-297, Editora Springer, 1995.

CRISTIANINI, N.; SHAWE-TAYLOR, J. **An Introduction to Support Vector Machine and Other Kernel-Based Learning Methods**. United Kingdom: Cambridge, 2000.

DESWAL, S.; PAL, M. **Artificial Neural Network based Modeling of Evaporation Losses in Reservoirs. World Academy of Science, Engineering and Technology**. Vol: 2. International Science Index, No:3, 2008.

DIAS, M. S. **O uso de máquina de suporte vetorial para regressão (SVR) na estimação da estrutura a termo da taxa de juros do Brasil**. Dissertação (Mestrado em Matemática) – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007.

ELIAS, R. M. S.; LIPENER, C.; URAS, R.; PAVÊS, L. Ceratocone: fatores prognósticos. **Arquivos Brasileiros de Oftalmologia**. São Paulo, v. 68, nº 4, pag. 491 – 494, Agosto, 2005.

FELICIANO DE DEUS, N. A.; **Estudo do endotélio corneal e suas mudanças em pacientes com ceratocone segundo o grau de progressão da ectasia** (Trabalho de conclusão de curso – Optometria). Universidade do Contestado, UnC, Canoinhas, Santa Catarina, 2015.

FERRÃO, M. F.; MELLO, C.; BORIM, A.; MARETTO, D. A.; POPPI, R. J. LS-SVM: uma nova ferramenta quimiométrica para regressão multivariada. Comparação de modelos de regressão LS-SVM e PLS na quantificação de adulterantes em leite em pó empregando NIR. **Revista Química Nova**. Vol. 30. Edição nº 4. São Paulo. Julho/Agosto 2007.

GEVERT, V. G. **Análise de crédito bancário com o uso de modelos de regressão logística, redes neurais e Support Vector Machine**. Dissertação de Mestrado (Métodos Numéricos em Engenharia). Universidade Federal do Paraná, 2009.

GUNN, S. R. **Support Vector Machines for Classification and Regression**. University of Southampton. Technical Report Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science, 1998.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation**. Prentice-Hall, New Jersey, 2 edition, 1999.

HEARST, M. A.; SCHÖLKOPF, B.; DUMAIS, S.; OSUNA, E.; PLATT, J. Trends and controversies - support vector machines. **IEEE Intelligent Systems**, 13(4):18–28, 1998.

HSU, C. W.; CHANG, C. C.; LIN, C. J. **A Practical Guide to Support Vector Classification**. Technical Report. Department of Computer Science, National Taiwan University, 2003.

JONHSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. 2nd ed., **Prentice-Hall Inc.**, New Jersey, 1988.

KECMAN, V. **Support Vector Machines – An Introduction**. The University of Auckland, School of Engineering, Auckland, New Zealand, 2005.

LAI, K. K.; YU, L.; ZHOU, L.; WANG, S. **Credit Risk Evalution with Least Square Support Vector Machine**. China: RSKT, 2006.

LIMA, C. A. de M. **Comitê de máquinas: uma abordagem unificada empregando maquinas de vetores-suporte**. Tese (Doutorado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2004.

LORENA, A. C.; CARVALHO, A. C. P. L. F. **Introdução às Máquinas de Vetores Suporte (Support Vector Machines)**. Relatórios técnicos do ICMC, Instituto de Ciências Matemáticas e de Computação, USP, 2003.

MAZUZE, A. N. F.; MENDES, R. T. **Mudanças refrativas em pós-operatório de pacientes submetidos à cirurgia de Pterígio pela técnica de transplante conjuntival** (Trabalho de conclusão de curso). CURSO DE OPTOMETRIA. UNIVERSIDADE DO CONTESTADO, UnC. CANOINHAS, SC. 2014.

MITCHELL T. **Machine Learning**. McGraw Hill, 1997.

MORALES, A. K.; SIMARI, G. R. *Advances in Artificial Intelligence*. **IBERAMIA 2010: 12th Ibero-American Conference on AI**, Bahía Blanca, Argentina, November 1 – 5 , 2010.

MOREIRA, Sally M.; MOREIRA, Luciane. Adaptação de lentes de contato em ceratocone. **Ver. Sinopse da Oftalmologia**. Ano 2 nº 1 São Paulo/SP 2000.

MORETTIN, P. A. & TOLOI, C. M. de C. **Previsão de Séries Temporais**. Editora Atual, São Paulo, SP, 1978.

MORETTIN, P. A. & TOLOI, C. M. de C. **Modelos para previsão de Séries Temporais**. IMPA, Rio de Janeiro, RJ, 1981.

MORETTIN, P. A. & TOLOI, C. M. de C. **Análise de Séries Temporais**. Editora Edgard Blücher, São Paulo, SP, 2006.

NETTO, A. V. System Based on Computational Intelligence for Ophthalmology Image Understanding. **IEEE LATIN AMERICA TRANSACTIONS**, VOL. 3, NO. 5, DECEMBER 2005.

ÖZGÜNDÜZ, E.; SENTÜRK, T.; KARSLIGIL, E. Off-Line Signature Verification and Recognition by Support Vector Machine. In: **13th European Signal Processing Conference**, Turkey, 2005.

PARRELLA, F. **Online Support Vector Regression**. Master Science Thesis. Department of Information Science, University of Genoa, Italy, 2007.

REDE de estações meteorológicas automáticas do INMET, Instituto Nacional de Meteorologia. Disponível em <http://www.portal.ufpr.br/tutoriais_normaliza/referencia_exemplo.pdf>. Acesso em: 30 de janeiro de 2016.

RUAS, G. I. S.; BRAGATTO, T. A. C.; LAMAR, M. V.; AOKI, A. R.; ROCCO, S. M. Previsão de Demanda de Energia Elétrica Utilizando Redes Neurais Artificiais e *Support Vector Regression*. In: **International Symposium on Communications, Control and Signal Processing**, 3. Piscataway, NJ : v. 1. p. 1431-1435, 2007.

SAMUI, P. Application of Least Square Support Vector Machine (LSSVM) for Determination of Evaporation Losses in Reservoirs. **Engineering**. Vol. 3, nº 4, 2011.

SANTOS, L. T. **Abordagem da máquina de vetor suporte otimizada por evolução diferencial aplicada à previsão de ventos**. Dissertação de Mestrado (Engenharia Elétrica). Universidade Federal do Paraná, 2013.

SAPANKEVYCH, N. I.; SANKAR, R. Time Series Prediction Using Support Vector Machines: A Survey. **IEEE Computational Intelligence Magazine**, University of South Florida, USA, 2009.

SHAH, R. S. **Least Squares Support Vector Machine**. 2005.

SMOLA, A. J.; BARKETT, P. SCHÖLKOPF, B.; SCHUURMANS, D. **Advances in Large Margin Classifiers**. MIT Press, 1999.

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, vol. 14. p. 199-222, 2004.

SOUZA, M. B. **Rede de aprendizado supervisionado como método de auxílio na detecção do ceratocone** (Tese de Doutorado em Oftalmologia). Faculdade de Medicina da Universidade de São Paulo, São Paulo, 2011.

SUYKENS, J. A. K.; VANDEWALLE, J. **Least squares support vector machine classifiers**. Neural Processing Letters, Springer Netherlands, v. 9, p.293 – 300, ISSN 1370-4621, 1999.

SUYKENS, J. A. K.; GESTEL, T. V.; BRABANTER, J. D.; MOOR, B. D.; VANDEWALLE, J. **Least Squares Support Vector Machines**. World Scientific Publishing Co. 2002.

TEIXEIRA, J. A. J. **Metodologia para implementação de um sistema de gestão de estoques baseado em previsão de demanda**. Dissertação de Mestrado Profissionalizante em Engenharia. Universidade Federal do Rio Grande do Sul, Escola de Engenharia, 2004.

TRAFALIS, T. B; INCE, H. Support Vector Machine for Regression and Applications to Financial Forecasting. **Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks**. Italy. v. 6. p. 348 – 353, 2000.

UOL. Economia Cotações. Disponível em: <<http://economia.uol.com.br/cotacoes>.> Acesso em 28 de fevereiro de 2016.

VAPNIK, V. N.; CHERVONENKIS, A. Y. **Theory of Pattern Recognition**. Nauka, Moscow, 1974.

VAPNIK, V. N., **The Nature of Statistical Learning Theory**. New York: Springer, 1999.

WOLD, H.O. **A Study in the Analysis of Stationary Time Series**. Stock. holm: Almqvist and Wiksel I. 1938.

YU, C.; LIU, Z.; MCKENNA,T.; REISNER, A. T. REIFMAN, J. A Method for Automatic Identification of Reliable Heart Rates Calculated from ECG and PPG Waveforms. **Journal of the American Medical Informatics Association**. Stanford, v.13, n.3, June 2006.

ZIEN, A.; RÄTSCH, G.; MIKA, S., SCHÖLKPF, B.; LENGAEUER, T.; MÜLLER, K. R. **Engineering support vector machine Kernels that recognize translation initiation sites in DNA**. Bioinformatics, 16:906–914, 2000.

**ANEXOS 1 – DADOS COLETADOS PELOS EXAMES PARA A DOENÇA
CERATOCONE (FELICIANO DE DEUS, 2015)**

VALORES PARA O GRAU 1

ID PX	OLHO	IDADE	SEXO	K CENTRAL	KISA%	GRAU	% (CV)	DENSIDADE	% HEX.
325	E	26	F	42,87	263%	I	31,7%	2699,2	70%
1774	D	26	M	47,11	131%	I	28,5%	2643,2	65%
286	E	26	M	46,74	324%	I	41,0%	2292,6	55%
2891	D	31	M	46,87	126%	I	29,7%	2805,2	69%
227	E	26	M	45,51	251%	I	34,4%	2407,7	33%
4352	E	28	F	45,34	541%	I	33,0%	2438,9	50%
4352	D	28	F	46,19	349%	I	33,1%	2296,7	54%
4648	E	35	M	47,15	213%	I	30,6%	2322,7	43%
4648	D	35	M	46,91	494%	I	32,9%	2227,9	65%
5232	D	34	M	45,28	160%	I	38,6%	2283,9	59%
408	D	30	F	47,84	558%	I	30,4%	2848,5	64%
2012	D	26	M	47,45	618%	I	33,9%	2603,4	71%
3705	E	31	M	45,44	930%	I	34,6%	3152,3	61%
3705	D	31	M	45,35	288%	I	31,6%	3107,0	45%
5020	D	23	M	45,24	264%	I	30,1%	2650,1	51%
1916	D	24	M	47,65	251%	I	34,2%	2673,8	55%
778	D	21	F	47,62	481%	I	28,4%	2773,8	81%
124	E	24	M	46,92	957%	I	31,5%	3082,2	77%
1240	D	23	M	43,76	404%	I	31,7%	2798,0	61%
880	E	23	M	47,79	875%	I	39,6%	2540,4	61%
343	D	24	F	43,51	315%	I	30,1%	3349,3	61%
2245	E	21	M	44,28	659%	I	34,6%	3299,4	63%
4468	E	27	M	45,44	355%	I	38,6%	3165,0	31%
4468	D	27	M	45,10	318%	I	35,8%	2711,9	42%

VALORES PARA O GRAU 2

ID PX	OLHO	IDADE	SEXO	K CENTRAL	KISA%	GRAU	% (CV)	DENSIDADE	% HEX.
5283	E	39	M	49,78	494%	II	34,5%	2702,2	56%
3539	D	23	F	52,16	4629%	II	29,0%	2800,7	43%
4636	E	36	M	49,21	904%	II	44,6%	3327,2	64%
1649	E	18	M	48,32	553%	II	35,0%	2809,3	60%
178	E	22	M	48,64	504%	II	30,0%	2356,2	69%
343	E	24	F	51,01	5249%	II	29,8%	3286,1	51%
1916	E	24	M	51,03	1332%	II	39,5%	2535,4	67%
2010	E	20	M	52,64	712%	II	29,6%	2820,0	83%
880	D	23	M	48,41	780%	II	32,4%	2666,8	64%
5261	D	32	F	49,06	400%	II	39,7%	2745,4	49%
7	D	36	F	48,92	346%	II	33,9%	2290,7	53%
227	D	26	M	49,04	820%	II	50,3%	2465,6	45%
2053	D	28	M	52,25	474%	II	50,3%	2302,6	29%
2053	E	28	M	50,69	601%	II	34,2%	3154,6	48%
286	D	26	M	48,88	468%	II	33,5%	2631,3	56%
5147	E	17	F	50,48	405%	II	38,5%	3179,1	39%
2725	E	33	F	50,25	3478%	II	35,3%	3013,3	44%
2725	D	33	F	49,28	1730%	II	29,8%	2521,9	54%
3506	E	30	M	52,41	8766%	II	33,9%	3263,7	65%
5534	D	29	M	49,8	433%	II	24,1%	2365,6	65%
316	D	29	F	49,13	533%	II	33,5%	2920,0	60%

VALORES PARA O GRAU 3

ID PX	OLHO	IDADE	SEXO	K CENTRAL	KISA%	GRAU	% (CV)	DENSIDADE	% HEX.
3396	E	22	M	54,24	7270%	III	29,1%	3025,0	80%
5487	E	28	M	55,91	4324%	III	39,2%	2696,1	22%
2052	D	28	M	53,36	11999%	III	28,3%	2604,5	50%
4804	D	29	M	54,74	7593%	III	36,5%	2643,9	65%

VALORES PARA O GRAU 4

ID PX	OLHO	IDADE	SEXO	K CENTRAL	KISA%	GRAU	% (CV)	DENSIDADE	% HEX.
4669	E	16	F	68,07	20346%	IV	30,8%	3008,7	65%
3517	E	40	F	58,08	18715%	IV	43,6%	3823,0	49%
2010	D	20	M	56,16	822%	IV	28,9%	2773,8	69%
124	D	24	M	57,17	8784%	IV	33,2%	2746,8	50%
3506	D	30	M	57,6	22940%	IV	35,9%	3557,6	61%

**ANEXO 2 – DADOS DO RESERVATÓRIO DE AGUA MANASGAON
(DESWAL, PAL, 2008)**

Mês	Semana	Evaporação (mm/dia)	Temperatura média do ar (° C)	Velocidade média do vento (m/s)	Horas de sol	Umidade relativa do ar
Janeiro	1	3,4	19,64	3,3	8,8	64,35
	2	3,3	20,29	2,9	8,3	61,35
	3	3,4	21,47	3,2	8,5	60,1
	4	3	20,86	3,8	6,9	60,45
Fevereiro	1	3,4	21,99	3,5	9	59,35
	2	3,9	23,59	3,8	8,9	55,7
	3	4,3	24,52	4,4	9,1	54,25
	4	4,7	25,62	4,1	8,9	48,4
Março	1	6,2	26,31	4,7	10,1	45,45
	2	5,6	26,89	4,9	8,6	45,1
	3	6,5	29,18	5	9,3	40,75
	4	5,9	30,04	5,4	8,2	40,2
Abril	1	8,1	31,52	5,8	10	42,55
	2	6,8	32,44	6,7	8,3	40,4
	3	8,8	33,36	6,4	9,8	38,05
	4	9,8	34,7	8	9,3	36,4
Maio	1	12,4	35,46	8,5	11,2	37,6
	2	11,9	34,85	9,5	10,2	44,4
	3	12,2	34,41	11,1	9,3	49,3
	4	11,2	34,68	10,3	8,3	46,6
Junho	1	15,3	34,05	10,7	9,4	53,3
	2	9,8	31,86	9,1	5	65,25
	3	7,5	30,07	10,4	4,1	74
	4	5,7	29,68	10,8	4,1	73,9
Julho	1	7,5	29,66	8,8	5,7	73
	2	5,7	29,01	9,3	3,9	77,6
	3	4,9	27,52	8,2	2,6	82,8
	4	3,3	27,37	8,1	3,4	83,95
Agosto	1	3,9	27,1	6,7	3,2	87,65
	2	3,1	26,7	6,8	2,5	86,9
	3	2,9	26,56	6,6	2,4	86,25
	4	3	26,88	6,1	3,7	84,85
Setembro	1	3,1	26,85	6,1	4	84,7
	2	3,3	27,43	5,4	5,6	79,3
	3	4,1	27,86	4,3	5,8	78,7
	4	4,4	28,58	2,9	6,3	78,95
Outubro	1	4,9	27,83	2,9	8	77,5
	2	4,3	27,43	2,8	7,3	75,65
	3	4	25,65	2,2	8,8	69,25
	4	3,5	24,62	2,5	7,6	65,8
Novembro	1	4,5	24,78	2,5	10,1	66
	2	3,5	24,47	2,5	7,4	69,35
	3	3,7	23,03	2,1	8,9	60,5
	4	3,4	22,47	2,2	8,7	58,5
Dezembro	1	3,7	20,78	2,4	9,6	61,4
	2	3,3	20,99	2,4	8,3	61,2
	3	3,3	20,21	2,5	8,1	61
	4	2,8	20,38	3,2	6,5	60,2

**ANEXO 3 – VALORES DOS EXAMES DE ASTIGMATISMO
(MAZUZE E MENDES, 2014)**

Paciente	CERATOMETRIA			REFRAÇÃO	
	MAIS PLANO	MAIS CURVO	ASTIG. CORNEANO	ESFERA	CILINDRICO
1	41,25	42,5	1,25	4,25	-1,75
2	43,75	47	3,25	5,25	-2
3	37,75	44,75	7	4,5	-4,25
4	43,5	44,25	0,75	0	-0,75
5	42,25	43,5	1,25	0,5	-0,25
6	43,25	45,5	2,25	0,25	-0,75
7	42	45,25	3,25	1	-1
8	43,25	45,75	2,5	1	-1
9	44,25	45,75	1,5	0,5	-0,25
10	37,75	44,5	6,75	3,25	-4
11	42	42,25	0,25	0,25	-0,5
12	42,75	44	1,25	0,5	-0,75
13	42	45,25	3,25	1	-1
14	43,25	44,25	1	1	0
15	43	44,25	1,25	-0,25	-0,5
16	41,25	44,75	3,5	4,25	-1,25
17	42,75	43	0,25	2,75	-0,5
18	44	45,25	1,25	0	0
19	44,25	44,5	0,25	0,25	-0,5
20	42,25	43,25	1	-0,5	-1,25
21	44	45,75	1,75	0,25	-0,5
22	42,5	43,25	0,75	0,5	-1
23	43,5	44,5	1	0,5	-0,75
24	42,75	43,25	0,5	2	-1,5
25	43,25	44,25	1	1	0
26	43,5	44,75	1,25	-0,75	0

**ANEXO 4 – VALORES CLIMÁTICOS PARA A PREVISÃO DA TEMPERATURA
(REDE, 2016)**

Temperatura	Umidade (%)	Ponto de orvalho (° C)	Pressão (hPa)	Vel. Vento (m/s)	Rajada de vento (m/s)	Radiação (Kj/M2)	Chuva (mm)
17,92619	87,583	15,68899	913,2833	2,085	5,292262	521,7119	28,4
16,78839	90,143	15,09048	912,0354	2,255	5,765476	411,0296	41,2
20,83304	81,565	17,20536	911,5848	1,977	5,266071	768,2757	28,4
19,8369	81,824	16,40268	907,183	2,498	6,09881	784,9756	59,2
18,79554	87,063	16,49583	909,3705	1,931	5,055357	555,4894	78,4
18,44583	87,301	16,09851	911,9339	1,623	4,504167	569,859	30,12
20,3875	80,744	16,70595	910,3051	2,246	5,664286	718,5862	50,0
22,48869	71,461	16,64524	911,1601	2,289	5,977976	975,2732	18,6
22,28929	82,137	18,76994	912,0104	1,976	5,301786	755,6606	138
22,01786	83,956	19,03839	909,3122	2,345	5,956548	607,8809	40
21,10744	79,318	17,18036	911,3182	2,263	5,747024	684,8435	17,8
21,54089	77,995	17,20938	908,5104	2,177	5,685938	797,8209	103,2